Probing Protein Interactions with Stapled Peptides: Myc Family and Insulin Receptor

Probing Protein Interactions with Stapled Peptides:

Myc Family and Insulin Receptor

A dissertation presented

by

Rebecca Yue Liang

to

The *Department of Chemistry and Chemical Biology*

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Chemistry

Harvard University
Cambridge, Massachusetts

May 2013

Probing Protein Interactions with Stapled Peptides:

Myc Family and Insulin Receptor

Abstract

One of the most exciting frontiers of expanding pharmacopeia to combat currently untreatable diseases is achieving specifically and potently disruption of unwanted protein-protein interactions where traditional small molecule drugs tend to fall short. Our laboratory has developed the methodology of peptide stapling and pioneered successful applications in multiple disease models since its induction over a decade ago. One common feature of past applications is the use of a single stapled peptide in helical form, derived from the natural binding interface of target proteins. This dissertation ventures into protein interactions that involve multiple components and sites and explores the extended use of stapled peptides in these volatile settings.

Up-regulation of the proto-oncogene Myc, a transcription factor in the Myc/Mad/Max network, is linked to abnormal growth and evasion of apoptosis in many types of cancers. Transcriptional activation by Myc roots in three major interactions: heterodimerization with Max, binding to DNA, and recruitment of histone modifying proteins. Max, the functional opposite of Myc, undergoes similar dimerization and DNA binding but solicits repression machinery that leads to gene silencing. The three distinct interactions – dimerization, DNA binding, and recruitment of transcriptional activators/repressors – are achieved by structural domains that have been shown to be portable by protein engineering and chimeric studies. We designed and synthesized

artificial transcriptional factors by fusing a stapled repression domain to a recombinant DNA-binding and dimerization domain via different methods. The physical properties of the fusion proteins have kept us from testing their potential in biochemical assays or *in vivo* models; nonetheless, we have gained experience with the molecules that may aid future efforts on improving their properties.

The insulin receptor (IR) ectodomain crystal structure shows the C-terminal domain of the α-subunit (α-CT) lying across the central α-sheet of L1 domain, indicating its critical role in insulin binding. Previously, surrogate peptides have been evolved by phage display to bind to one of the two insulin-binding sites on IR. Among these hits is S371, a prototypical site 1 peptide with sequence similarity to α-CT of IR, possibly a mimetic. We synthesized a comprehensive library of stapled peptides based on S371 including all possible locations of hydrocarbon crosslink along the sequence. Screening of the library for IR and Akt1 phosphorylation has identified several potent antagonists of IR signaling, with IC50 at about 1 μM. All hits feature non-natural amino acid substitutions on fixed "hot spots" on the S371 sequence that render the crosslink on just one face of the helix, presumably to preserve the binding interface. Biochemical assays were performed to support the structural model of S371 occupying the same binding site as αCT. We expect further optimization of the stapled peptide hits to achieve higher potency and agonistic effects, and aim for solving the structure of receptor: peptide complex that will help unravel mechanisms of receptor activation.

Table of Contents

Acknowledgements.....vi

List of Figures.....viii

List of Abbreviations.....x

Chapter I – Design and Synthesis of Stapled Peptides and Proteins for Targeting the Myc/Mad/Max Network.....1

Chapter II – Experimental Optimization of Semi-synthetic Transcriptional Factors.....32

Chapter III – Design, Synthesis, and Screening of a Library of Stapled Insulin Receptor Binding (SIRB) Peptides.....48

Chapter IV – Evaluation of Active SIRB Peptides: Binding Site, Specificity, and Potency Optimization.....73

Acknowledgements

I want to start this thesis by thanking my advisor, Professor Gregory Verdine, on a few accounts. First, Greg's genuine passion and unorthodox approach for tackling difficult problems have always been quite infectious, while his generous patience and support had stopped me from derailing more than once. Second, I have learned tremendously by listening to Greg – not only because he is knowledgeable and artful with speech, but also from the high standards he sets for us: be thoroughly scientific and methodological in designing research, be effective and bold in communication and presentations, pay attention to aesthetic details, etc. These standards, some of which I am still trying to reach, will certainly make a lasting impact on my future careers. Third, I thank Greg for showing me that a scientist could capitalize on his or her expertise outside of academia, which in my opinion is an audacious and inspirational stance not all PhD advisors would be able to take.

I thank Professors Stuart Schreiber and Dane Wittrup, for serving as my graduate advisory committee (GAC) members. They have always given me constructive feedback, new and alternate perspectives on problems that had begun to feel "old", and plenty of moral support during difficult times. Prof. Schreiber had helped me connect with experts and resources at the Broad Institute many times, and there are on-going joint efforts that we expect to lead to new insights.

In the laboratory, I have been fortunate to work with and learn from amazing individuals. I thank Dr. Minyun Zhou, for a wonderful collaboration near the end of my tenure on the insulin receptor project. Her proficient skills and hard work helped us obtain the protein that had been the bottleneck of many experiments. I thank Matt Kolar, a great colleague and friend, who had worked with me to make the SIRB library and without whom

the synthetic expedition would have been quite overwhelming. I want to thank Dr. Jerry Hilinksi, for teaching me practically everything about peptide chemistry from scratch when I first started and bearing through my continuous consultations with good humor in later years. I thank the newly minted Dr. John McGee, an exceptional labmate and classmate of my own year, for providing useful comments and suggestions on my projects and for setting up the example of good lab citizenship. I would also like to thank our administrators Lydia Carmosino and Betsy Donovan (Betsy had been there for my first two years in the lab). They had taken care of almost everything else so I could focus on the research.

I consider myself most fortunate to have finished graduate school side-by-side with a best friend. I have known Jing Yang since our first days at Harvard five years ago, and her companionship has made these long years full of challenges, setbacks, and emotional turmoil so much more endurable.

I thank my boyfriend, Billy Au, for always being there for me even when he is half a globe away. Having graduated from the same department a year ago, he had been a great role model and source of moral support. His love and support had not only meant happiness to me, but also peace of mind, enabling me to traverse through the unknowns and uncertainties of the PhD journey with ease.

Finally, I want to thank my parents, Jing Wang and Jian Liang. There is no clever combination of words that can adequately express how much their unconditional love and blind faith means to me, nor does sheer volume of text do my gratitude much justice. I will simply say this: everything I am, everything I have accomplished, I owe to my parents. They will continue to be my biggest source of inspiration for higher impact in the ensuing chapters of my life.

List of Figures

Chapter I

Figure 1.1 – Domain organization of Myc family members……5

Figure 1.2 – Structures of Myc family dimers……8

Figure 1.3 – Schematic representation of semi-synthetic transcription factor……9

Figure 1.4 – Synthesis of transcription factor via native chemical ligation……11

Figure 1.5 – Interaction between Mad and Sin3 ……12

Figure 1.6 – Sequences of wild-type SID and stapled repression domain (SRD) peptides……13

Figure 1.7 – Cellular uptake of SRD peptides. ……14

Figure 1.8 – Measurements of affinity of SRD peptides to Sin3. ……15

Figure 1.9 – Synthetic scheme for C-terminal thioester peptide to use in NCL……17

Figure 1.10 – Covalent modification of N-terminal cysteine on Max bHLH-Zip……19

Figure 1.11 – Native chemical ligation of wtSID and Max bHLH-Zip…… 20

Figure 1.12 – Product and by-product formation in NCL…… 21

Chapter II

Figure 2.1 – Proposed synthetic scheme of cysteine-maleimide conjugation……33

Figure 2.2 – "One-pot" conjugation of SRD and Max……34

Figure 2.3 – LC/MS analysis of conjugation and subsequent purification……35

Figure 2.4 – Second-generation SRD peptides aimed to increase solubility……37

Figure 2.5 – Measurement of binding affinity of new SRD peptides to Sin3……39

Figure 2.6 – Removal of DNA from recombinant Max and de-aggregation……40

Figure 2.7 – Denaturation and refolding of Max bHLH-Zip……42

Chapter III

Figure 3.1 – Representative domain structure of the insulin receptor....51

Figure 3.2 – Structure of the a-CT segment of IR and sequence comparison with S371....52

Figure 3.3 – Schematic representation of all types of hydrocarbon cross-link implemented on peptide S371....55

Figure 3.4 – SIRB antagonism and agonism....58

Figure 3.5 – Validation of SIRB peptides....59

Figure 3.6 – Dose-response of antagonist SIRB-B5....61

Figure 3.7 – Secondary structure of S371 and select SIRB hits....61

Figure 3.8 – Summary of active SIRB peptides....63

Chapter IV

Figure 4.1 – Use of photo-activated cross-linking (PCL) as a means to reveal binding site of a ligand's binding site on target protein....75

Figure 4.2 – Incorporation of BPA into SIRB-B5....77

Figure 4.3 – IR ectodomain constructs....79

Figure 4.4 – Photo-activated cross-linking of SIRB-B5-BPA and L1-CR-L2....82

Figure 4.5 – Loss of function of SIRB-B5 BPA peptides....83

Figure 4.6 – Affinity pull-down of IR by SIRB-B5....85

Figure 4.7 – Design of competition pull-down to probe the binding site of SIRB-B5....86

Figure 4.8 – Results of competitive pull-down assay....87

Figure 4.9 – Comparison of Site 1 between IR and IGF-1R....89

Figure 4.10 – SIRB-B5 antagonism on IGF-1R....91

Figure 4.11 – Design of SIRB homodimer synthesis....94

Figure 4.12 – Increase of agonist potency by dimerization of SIRB-D2....95

Figure 4.13 – Change in position of α-CT during insulin binding....97

List of Abbreviations

bHLH-LZ	basic helix-loop-helix/leucine zipper
BPA	Benzoylphenylalanin
CD	circular dichroism
CML	chronic myeloid leukemia
CR	cysteine-rich domains
Dbz	diaminobenzoic acid
DCE	1,2-dichloroethane
DCM	dichloromethane
DIPEA	N,N-diisopropylethylamine
DMEM	Dulbecco's Modified Eagle Media
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
E-Box	enhance box
EGFR	epidermal growth factor receptor
ELISA	enzyme-linked immunosorbent assay
EPO	erythropoietin
ER	estrogen receptor
FITC	fluorescein isothiolate cyanide
Fmoc	N-a-Fluorenylmethoxycarbonyl
Fn	fibronetcin type III domains
FP	fluorescence polarization
GPCR	G-protein coupled receptors
GST	Glutathione S-transferase
HATU	*O*-(7-azabenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate
HBS	hydrogen bond surrogate
HCTU	2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HDAC	histone deacetylase
HOBt	Hydroxybenzotriazole
HPLC	high-performance liquid chromatography
IC50	half maximal inhibitory concentration
IGF-1R	Type-1 insulin-like growth factor receptor
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	insulin receptor
IRS	insulin receptor substrate
K_d	dissociation constant
L1, L2	leucine rich repeats
LC/MS	liquid chromatograpy/mass spectrometry

LIC	ligation-independent cloning
MALDI	matrix assisted laser deionization
Mesna	sodium 2-sulfanylethanesulfonate
Nbz	N-acyl-benzimidazolinone
NCL	native chemical ligation
Ni-NTA	nickel-nitriloacetic acid
NMP	*N*-methyl-2-pyrrolidone
PAH	paired amphipathic helix
PAP	*para*-azido-phenylalanine
PCL	photo-activated cross-linking
PI3K	phosphatidylinositol 3'-kinase
PKB/Akt	protein kinase B
PyClock	6-chloro-benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate
RTK	receptor tyrosine kinases
SIRB	Stapled Insulin Receptor Binding Peptide
SMCC	Succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate
SPPS	solid phase peptide synthesis
SRD	stapled repression domain
TAD	transcriptional activation domain
TCEP	tris(2-carboxyethyl)phosphine
TF	transcriptional factors
TFA	trifluoroacetic acid
TIS	triisopropylsilane
TK	tyrosine kinase
TM	transmembrane
TPO	thrombopoietin
TRRAP	transcription/transformation domain-associated protein
α-CT	C-terminal a-subunit peptide

Chapter I –Stapled Peptides and Proteins for Targeting the Myc Network

Introduction

At the core of the modern approach for drugging a disease lie two doctrines of practice: first, find the genes and proteins whose aberrant behavior is the cause of the disease; second, find a way to either correct for or remove such behavior, frequently in the form of a chemical compound physically binding to the target and rendering it inactive. Though many breakthroughs in drug discovery have been made possible by efforts following these guidelines, the current situation is at its most challenging of all time on both regards.

For one thing, finding the right target(s) is hard. Long gone was the naïve belief that any disease could be pinned to one (or even just a defined number of) molecular deregulation(s) as its cause. The handful of single gene disorders are likely more of an exception than a rule, and have been better off used as poster-child stories to fill up introductory genetics or drug design text books that lure future generation of researchers or physicians. Most diseases that still elude effective treatment today have been found over long periods of research to implicate tens or hundreds of genes and proteins with no set pattern but large individual variation. Even in the rare and marvelous event when targeting of a single enzyme is sufficient to overcome disease state, such as the famous case of imatinib and its target BCR-

ABL in chronic myeloid leukemia (CML), resistance arises when the specific drug-protein interaction is compromised by mutations as seen in many CML patients under treatment og imatinib[1]. Indeed, the more knowledge we gain, the clearer it drives home that the battlefront of identifying and targeting the intricate molecular networks that cause diseases is mercurial and ever expanding.

For another, chemists strive to expand the reservoir of compounds and synthetic methodologies to yield potential therapeutics, but find it harder and harder to keep up with the growing complexity of biological tasks at hand. Most drugs in the pharmacopoeia today work in the same way – they bind to defined clefts in target proteins to shut them off functionally. Not surprisingly the list of targets druggable by such means also consists of similarly structured proteins; for example, more than a quarter are of the class of G-protein-coupled receptors (GPCR), which in turn represent less than 2% of the entire human genome[2]. Granted, the conventional way of drug discovery – screening of tens of thousands of small molecules for the ability of inhibiting a particular protein – has received tremendous success in presenting us with almost all of the drugs that became available in the last century, more than all previous ones combined. It turns out, however, that there are only so many low-hanging fruits available for picking, so to speak. As mentioned above, disease biology is increasingly found to involve complex processes such as multi-component interactions, crosstalk between proteins, DNA, and regulatory RNA, and global-level events that change cell fate altogether. Turning a single molecule off simply would not suffice, but beyond this realm screening of a traditional small molecule library does not have much to offer. Modulation of multifaceted networks and events requires a matching increase in sophistication and dimensionality of chemicals and methods in the toolbox. Facing this demand, chemists have stepped up to meet the challenge, developing new paradigms of synthetic strategy to yield structurally diverse and complex molecules in a

high throughput and scalable manner[3]. Much progress has been made on targeting undruggable interactions by screening new generations of libraries. In addition to small molecules, "big molecules", i. e. biologics have entered the arena. These are peptides, monoclonal antibodies, nucleic acids, etc. derived from relevant biological context and thus inherently better poised at targeting said context. But formulation and delivery of these more volatile molecules pose new problems unfamiliar to conventional medicinal chemistry. Nonetheless, biologics often exhibit insurmountable specificity and potency if their pharmacologic properties are optimized, and represent a new hope and foundation for the development of next-generation therapeutics.

Albeit daunting for the reasons discussed above and more, research in both doctrines continues to evolve and thrive. Revolutionary studies hit major periodicals every so frequently, and there has never before been a more exciting time to be part of the broadening community of disease-oriented biologists and chemists alike. The theme of this dissertation is precisely methodology and compound development with the aim of studying and potentially drugging perhaps some of the most invincible of all diseases: cancer and diabetes. Incidentally, our approach is one inspired by both small molecules and biologics: chemically-modified peptides.

The idea of stabilization of peptides in the α-helical structure through chemical modification is not new. Demand for methodologies that enable short peptides to retain their native α-helical conformation even when taken apart from their parent protein is ever growing, especially as helices account for much of the known protein-protein interfaces. Approaches in the past include cross-linking of side chains, hydrogen bond surrogate (HBS) helices, β-peptides and other peptidomimetics[4-7]. Although these methods yielded stabilized peptides or mimetics that demonstrated increased helical character *in vitro*, their use as probes or

potential therapeutics in disease biology has been scarce due to a variety of limitations, such as susceptibility to disruption or degradation in the physiological chemistry environment, loss of binding affinity to protein partners, low cell penetration level, etc.

Our laboratory has developed another way to "lock" peptides in their helical conformation via incorporation of non-natural amino acids: peptide stapling[8]. All-hydrocarbon cross-linking of a peptide via metathesis of olefin-containing unnatural amino acids, a technique developed in our lab, enables solid phase peptide synthesis (SPPS) of α-helices that exhibit enhanced helicity, cell-permeability, and metabolic stability. Significantly, stapled peptides have been shown to be exceptionally specific and highly potent to unwanted protein-protein interactions in multiple disease models both *in vitro* and *in vivo*[9-13]. Recent progress in our lab indicates that they may even transport various types of cargo attached to them into cells. Equipped with this powerful toolkit, we would like to undertake a fresh approach in tackling an omnipresent yet extremely elusive target in cancer biology: the Myc/Mad/Max transcriptional network.

The Myc/Mad/Max Transcription Factor Family

Proteins Myc and Mad belong to a family of eukaryotic transcriptional factors (TF) that possess the basic helix-loop-helix/leucine zipper (bHLH-LZ) motif. The domain organization of major members is shown in figure 1.1. Both of the proteins form heterodimers with Max, another bHLH-LZ TF, via interaction between the LZ domains. When dimerized, the bHLH domains of Myc and Mad recognize the same DNA sequence known as the E-Box[14-16]. Despite the numerous common features they share, Myc and Mad usually trigger opposite effects in gene expression. Simply put, Myc tends to activate transcription by association of

its N-terminus with TRRAP, a member of the HAT complex SAGA. Conversely, the N-terminus of Mad recruits Sin3, part of an HDAC complex, thus resulting in gene silencing[17-19].

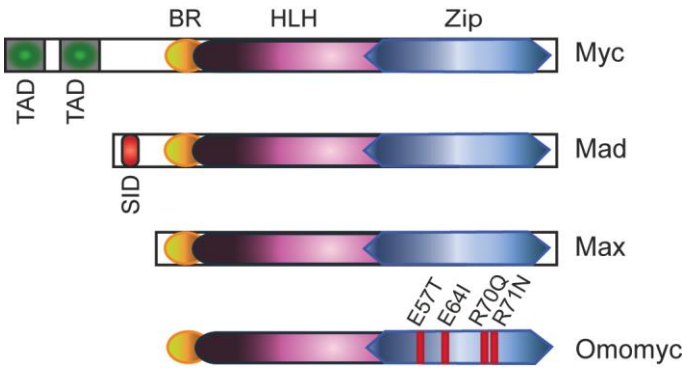


Figure 1.1. Domain organization of Myc family members. BR, basic region; HLH, helix-loop-helix domain; Zip, leucine zipper; TAD, transcriptional activation domain; SID, Sin3-interaction domain. Omomyc is derived from Myc by N- and C-terminal truncation down to the core bHLH-Zip domain, plus introduction of the four denoted amino acid substitutions.

Despite the direct involvement of its constitutive activity in many types of cancer, Myc has received far less attention as a potential drug target than it deserves. Firstly, many of Myc's target genes are essential for cell growth and maintenance. Secondly, disruption of protein-protein interactions is in general hard to achieve by small molecule drugs. Recent studies, however, have shown that a dominant interfering mutant, termed "Omomyc", inhibits lung tumor growth in mice, with tolerable and reversible effects on other regenerating tissues[20,21]. Encouraged by the therapeutic potentials of Myc, the propose study seeks to answer the other side of the question – whether Myc can be inhibited by methods not involving exogenous gene delivery.

Design of Stapled Repression Domains and Semi-synthetic Transcription Factors

Reviewing of successful applications of the stapled peptide strategy for targeting disease-causing, deregulated protein interactions (mostly oncogenic proteins) in the past shows a common feature of design: all stapled peptides developed for a host of different targets are based on sequences from the respective natural interaction partner; they form part of the binding interface that researchers were trying to disrupt. In this way, they could be

classified as dominant negatives that occupy the same site on the target protein as the "parent" molecule from which they are derived, blocking its access. By virtue of dramatic increases in binding affinity, stability, and bio-availability from installation of hydrocarbon staples, these dominant negatives become great competitive inhibitors that work in the same type of logic as small molecule drugs of the classical "bump-hole" strategy, except the "bump" and "hole" in question are much larger and flatter surfaces. While there is no doubt that this approach has the advantage of simplicity and can work well in certain systems, it is by far not a universal solution. Not all protein-protein interactions are as simple as the form of a single helix on one protein engrossed by a cleft on another. Many include more than site of interaction constituted by more than two components, as is the case of Myc/Max, Mad/Max dimers and their binding to DNA we discussed earlier.

Our lab had put considerable efforts on targeting the Myc/Max transcriptional machinery via a similar design of developing stapled peptides as dominant negatives. At first glance, the heterodimer of Myc and Max bound to DNA, although elaborate in structure and interaction motifs, still seemed quite amenable to the old approach. In theory, a peptide need not touch all of the interaction sites to disrupt formation of the complex – one is likely enough. The real conundrum lies in an underlying property of the multi-component system held together by interactions at different sites synergistically: any one of these sites alone is almost certainly characterized by relatively low binding affinity, nor is it obliged to have evolved high specificity by nature.

If the starting point is too low, incremental improvement on affinity by stapling does not guarantee the outcome of a particularly active compound that blocks native interactions by binding to just one site. In the Myc/Max/DNA system, numerous stapled peptides have been made that target two distinct interfaces, one at the leucine zipper responsible for

dimerization of proteins and the other at the basic region for protein-DNA binding. Although some hits were identified that are able to inhibit complex formation in *in vitro* assays, no peptide potent and specific enough to induce specific *in vivo* reprogramming has emerged from either group. In retrospect, a plausible explanation may be that the "native binding force", so to speak, is not only made up from multiple sites, but even within each site is spread over a large interaction surface – over the long (>40 amino acids) stretch of leucine zipper, for example, of the dimerization interface. When taken out of context and truncated to a peptide shorter than 25 amino acids, it is hard for the minimized helix to recapture all of native binding force. If the hope had been that perhaps the staple would compensate for the lost affinity, we now know better. The moral is, each system is different, and only empirical data could answer the question.

So how could we modify the strategy to better target systems that do not respond well to inhibition by a single helix? This question direly needs answering for the phenomenon is too omniscient for any researcher to ignore. Other than transcription factors, many of which regulate gene expression by having similar domain organization and structural motifs as Myc/Mad/Max, most cell-surface receptors and their respective ligand hormones also demonstrate the same multi-site binding behavior in their own way. The challenge is both intimidating and intellectually stimulating, and marks the central cause of this dissertation. The first two chapters describe new attempts at tackling the Myc/Mad/Max network, and –- the latter ones at the insulin: insulin receptor system.

Let us first return to the domain organization and structural biology of Myc/Max and Mad/Max (Figure 1.2). X-ray crystallography structures of Myc/Max and Mad/Max heterodimers in complex with DNA were solved by Nair and colleagues in 2003[22]. Both complexes crystallized were formed by truncated proteins containing the bHLH-Zip

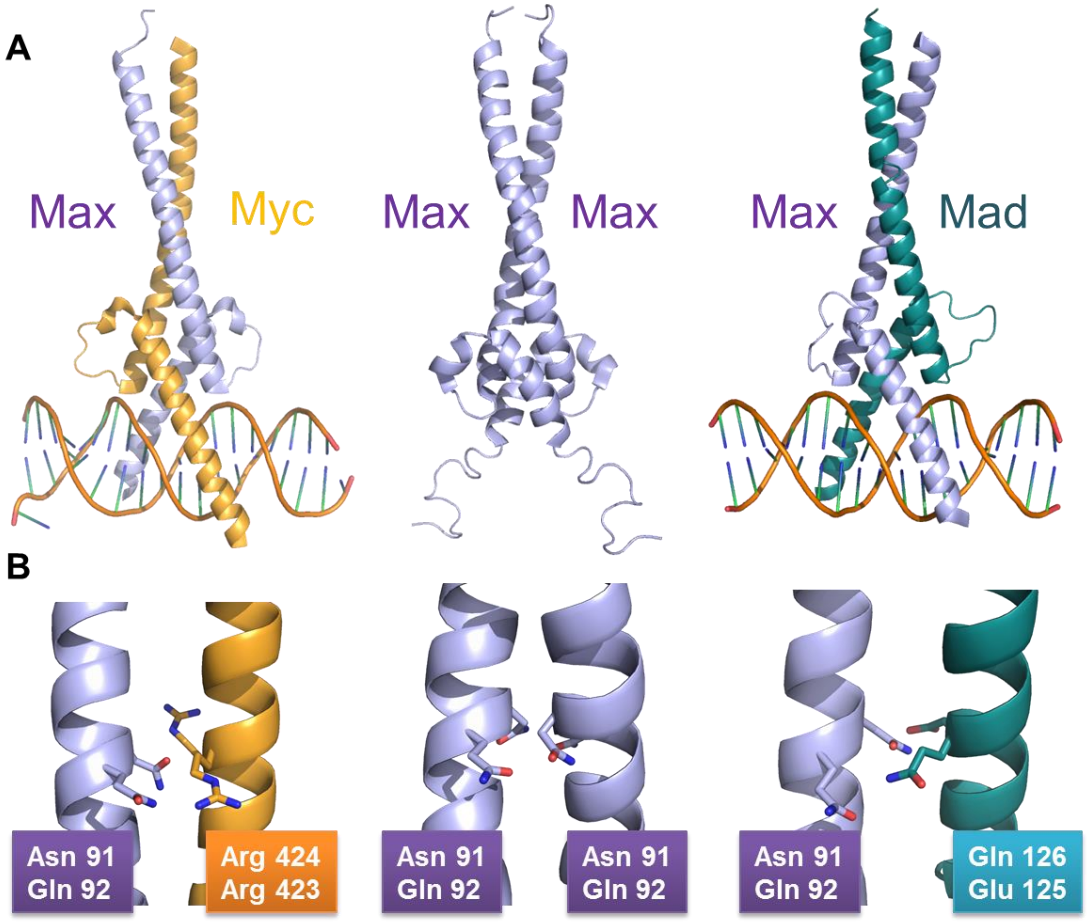


Figure 1.2. Structures of Myc family dimers. (A) Crystal structures of heterodimers Myc/Max and Mad/Max bound to Ebox DNA and solution NMR structure of Max/Max homodimer, PDB codes: 1NKP, 1NLW, 1R05, respectively. (B) Close-up view of helix 2/leucine zipper where dimerization specificity is governed by the combination of amino acid tetrad, side chains shown in sticks. Max alone is able to dimerize with both Myc and Mad, as well as homodimerize, via electrostatic interactions, whereas homodimerization of Myc or Mad would result in electrostatic repulsion at the tetrad, and is thus never observed.

domains only, which proved sufficient for DNA-binding and dimerization. Specifically, Ebox recognition and binding occurs at the basic region achieved mostly by electrostatic interactions between the four arginines and lysines of the protein and DNA bases. Dimerization happens at helix 2 and leucine zipper regions, and the interaction is less well understood. Taking these data together with the NMR solution structure of Max homodimer[23], It has been suggested that the specificity for heterodimer is governed by a tetrad of amino acids – Asn91 and Gln92 of Max, Arg423 and Arg424 of Myc, and Glu125

and Gln126 of Mad. Such combinations explain why the combinations of Max homodimer, Myc-Max and Mad-Max are possible, and that neither Myc nor Mad can homodimerize, as it would introduce electrostatic repulsion at the tetrad.

In addition to providing insights on binding and specificity, these structures of truncated proteins also allow us to consider new means for inhibiting native complex formation. Since we do not need to include more than the bHLH-Zip domains to completely reconstitute dimerization and DNA binding, could we somehow obtain such a minimized construct that is within synthetic or semi-synthetic reach, to use as a super dominant negative? If there is a hydrocarbon staple somewhere to stabilize the artificial bHLH-Zip and help deliver it into cells, it should be able to compete with Myc/Max in binding to EBox sequence. Such a construct would surely be free of shortcomings in low affinity or specificity suffered by shorter helices mentioned earlier.

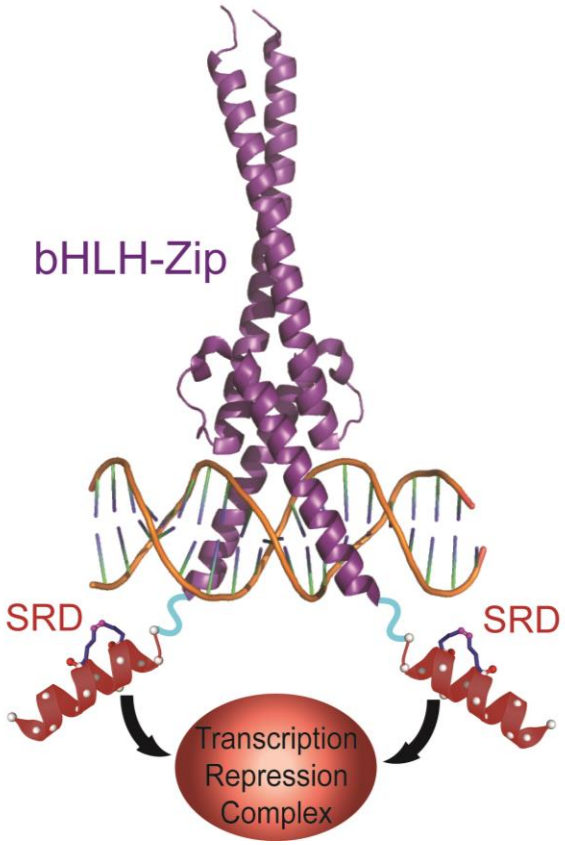


Figure 1.3. Schematic representation of semi-synthetic transcription factor. The proposed synthetic transcription factor aims to combat Myc deregulation via competing for its DNA binding site (Ebox) and recruiting repression machinery. It consists of a recombinant bHLH-Zip portion derived from Max and a synthetic portion – stapled repression domain (SRD).

Better still, what if we staple not the bHLH-Zip core, but rather a free-standing repression domain such as one derived from MAD – the Sin3 interacting domain or SID, and then fuse the stapled repression domain to bHLH-Zip? The resulting molecule will not only be a dominant negative, but also an active silencer of Myc-induced up-regulation. Rather than simply competing for DNA binding, it would possess another business end that works in the opposite direction to Myc. This ambitious vision is too enticing not to try.

Our design strategy for an artificial transcription factor with the aim of antagonizing Myc is illustrated in Figure 1.3. First, we will select a suitable minimized bHLH-Zip domain that retains Ebox-binding and dimerization capabilities from the Myc family members. In parallel, a stapled peptide version of a repression domain (SRD) will be developed that achieves high ligand binding affinity and cell penetration power. The two independently obtained entities will then be covalently linked. We expect the fusion product to inherit the many functions of the two building blocks. First, the hydrocarbon stapled region would enable cellular uptake of the entire fusion and eventually nuclear localization. Once nuclear, the bHLH-Zip would either self-dimerize or exchange to form heterodimers with Myc family members, and subsequently recognize and bind to Ebox DNA sequence in a competitive manner to Myc/Max. The SRD would then bind to its target, Sin3, which would in turn serve as a scaffold for recruitment of other transcriptional co-repressors. In summary, the biological outcome is expected to be down-regulation of Myc-dependent gene transcription by an SRD-fused reprogrammer.

As for the choice of the bHLH-Zip component, we first propose the use of the minimized Max bHLH-Zip since it has a short sequence and a reasonable track record in crystallography, and hence reasonable manageability, but there are other suitable candidates such as Omomyc or Mad itself. Once both components are ready, we would adapt a semi-synthetic

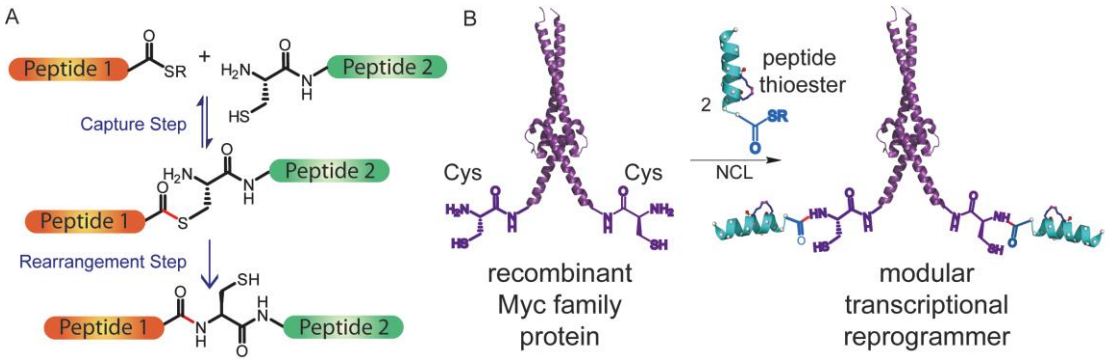


Figure 1.4. Synthesis of transcription factor via native chemical ligation. (A) Native chemical ligation entails the coupling of a peptide or protein having a C-terminal thioester with a peptide or protein having an N-terminal Cys residue, to produce a ligated product having a native peptide bond at the linkage site. (B) Ligation of an SRD with a C-terminal thioester onto a Myc family bHLH-Zip domain in which a cysteine is engineered at the N-terminus.

approach that connects the synthetic SRD peptide to the recombinant bHLH-Zip domain (Figure 1.4). A reaction scheme known as native chemical ligation (NCL)[24,25] will be employed for the present purposes, as this reaction generates a native peptide bond at the ligation site and has been used extensively in protein semi-synthesis[26-30] and our laboratory reported the first example of native chemical ligation on a recombinant protein[31]. The method of NCL yields a product that is "scarless" of any non-peptido functionalities and thus the closest imitation of a natural protein. For this reason it would be our top method of choice, the execution of which would entail specific requirements on peptide and protein chemistry. The N-terminal component, which in our case is the SRD, would need to have a thioester group on its C-terminus. The C-terminal fragment, the bHLH-Zip, has to begin with an N-terminal cysteine. The design and methods for installing such chemistry will be discussed in detail later in this chapter.

Synthesis and Evaluation of Stapled Repression Domain Peptides

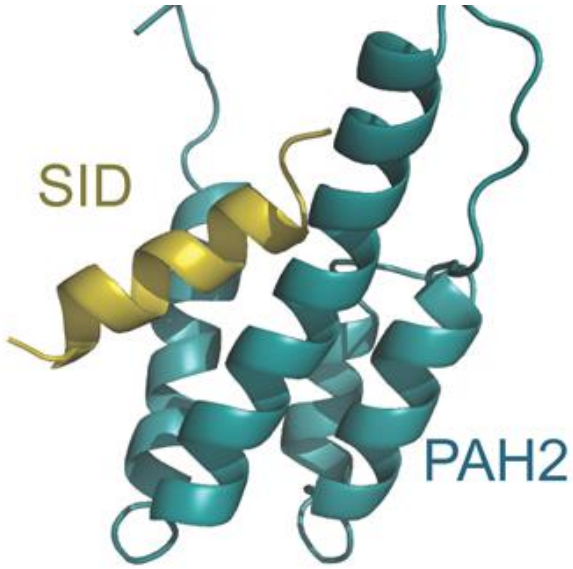


Figure 1.5. Interaction between Mad and Sin3 (solution NMR Structure). The Sin3-interaction domain (SID) from Mad (gold) bound to its target, the PAH2 domain of Sin3. PDB accession code: 1G1E.

Our first task was to obtain an active stapled repression domain that binds to Sin3 with high affinity and specificity, and preferably penetrates cells efficiently as it would have to carry cargo later. We examine the interaction between our SRD of choice and the transcriptional co-repressor it recruits in more detail (figure 1.5)[32]. The paired amphipathic helix (PAH) Domain of Sin3 is a sequence around 100 residues long that is highly conserved across species and structurally forms a left-handed four-helix bundle. It has been shown that the PAH domain alone is both necessary and sufficient for interaction Mad. Between helices α1 and α2 is a hydrophobic cavern where Sin3-interacting domain of Mad binds via its apolar face. SID in turn is also a highly conserved amphipathic helix containing a unique sequence motif depicted in figure 1.5B. It locates at the N-terminal of Mad, and the exactly length is debatable as both a short and long version differing in the inclusion of the C-terminal residues have been reported. Most of the residue interactions between SID and PAH are of hydrophobic nature, though notably the conserved glutamate near the C-terminal end of SID, the only charged residue in the sequence motif, engages ionic interactions with two lysines on Sin3. Much to our advantage, the natural SID peptide binds to the PAH of Sin3 quite tightly; the K_d of the full domain (24 amino acids) is about 100nM, a very good starting

point. Hence our primary goal in designing SRD was to maximize cellular uptake while preserving and perhaps further enhancing binding affinity.

Design of the initial batch of stapled peptides was under the principle of sampling as wide a range of staple positions throughout the sequence as possible while preserving the integrity of the binding face of the helix. Specifically in this case most of the conserved residues of the sequence motif were also left untouched (Figure 1.6). In some of the peptides we also mutated a few residues near the N-terminus to arginine, capped by a glutamate at the terminus, as it has been our experience with a large number of stapled peptides that a negative N-terminus and positive net charge together tend to favor greater helicity and cell permeation (unpublished data). Finally, there was one type of hydrocarbon cross-link – i, i+4 that we used for all peptides, partly for simplicity, partly for the already hydrophobic sequence that may become too greasy to work with using longer staples.

Following solid phase synthesis of the primary sequence via standard Fmoc chemistry, ring-closing metathesis was performed to close the macrocycle. To obtain compounds suitable for microscopy and binding assays, we appended fluorescein isothiolate cyanide (FITC) to

Compound	Sequence																									
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
wtSID			V	R	M	N	I	Q	M	L	L	E	A	A	D	Y	L	E	R	R	E	R	E	A	E	H
Min. Motif				X	X	X	Φ	Z	Z	Φ	Φ	X	A	A	X	X	Φ	E	X							
SRD1	βAla-	E R	L	R	R	R	I	*	M	L	L	*	A	A	N	Y	L	E	R							
SRD2	βAla-		V	R	R	R	I	*	M	L	L	*	A	A	N	Y	L	E	R							
SRD3	βAla-		V	R	R	R	I	Q	R	L	L	*	A	A	N	*	L	E	R							
SRD4	βAla-		V	R	M	N	I	Q	M	L	L	Q	A	A	N	R	*	E	R	R	*	R				
SRD5	βAla-		V	R	R	R	I	Q	M	L	L	E	A	A	N	Y	*	E	R	R	*	R				
SRD6	βAla-		V	R	M	N	I	Q	M	L	L	Q	A	A	N	R	L	E	R	R	*	R	E	A	*	H
SRDN	βAla-		V	R	R	R	I	*	M	L	L	*	L	A	N	Y	L	E	R							

Figure 1.6. Sequences of wild-type SID and stapled repression domain (SRD) peptides. wtSID was made to include the full-length canonical SID sequence. SRD1, 2, 3, and 5 were mutated in the N-terminal region to contain more arginine residues as an effort to increase cell penetration. SRDN was a point mutant of the identified hit SRD2 and used in binding assays as a negative control.

the N-terminus of the peptide after a β-alanine linker. (Here it is worth noting that it is not possible to N-terminally label of FITC directly after the last residue in the sequence; some form of linker not in the form of a natural amino acid is necessary lest side reactions occur. An interesting paper in 2009 addressed the issue, published just as I was making futile attempts at FITC-labeling without β-alanine[33]).

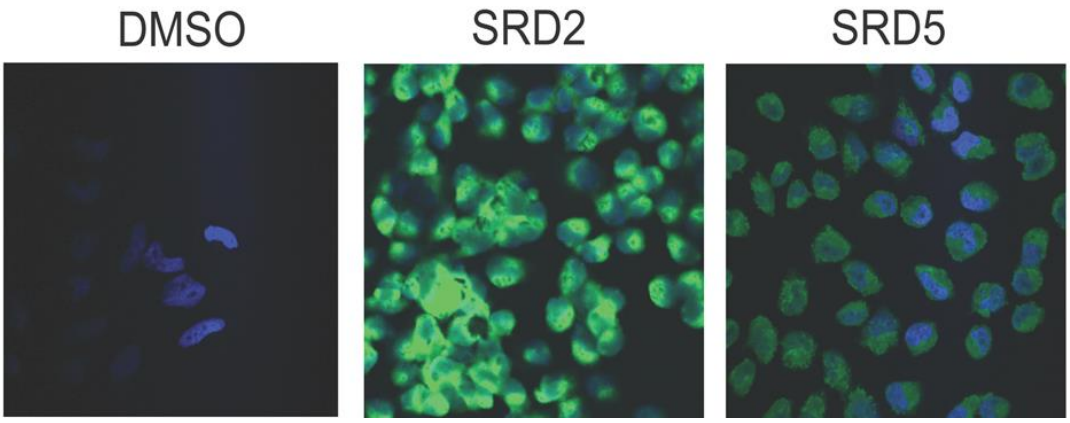


Figure 1.7. Cellular uptake of SRD peptides. HeLa cells were treated with FITC-conjugated SRD peptides for 4h at 37°C in the presence of serum. Confocal microscopy revealed robust cellular penetration of SRD2 and SRD5.

After synthesis was complete, peptides were purified by reverse-phase high-performance liquid chromatography (HPLC) and ready for assays. Cellular uptake of the peptides was assessed by treating HeLa cells with compounds for various durations followed by confocal microscopy. We found the most of these SRD peptides are capable of entering cells at 10μM to give quite strong fluorescent signals; representative images are shown in Figure 1.7.

Satisfied with the cell-penetrating power of SRD peptides, we proceeded to characterize their affinity to Sin3. The PAH2 domain of Sin3 was cloned and expressed from a pET24a vector that was kindly bestowed to us by Professor Radhakarishnan's lab. We performed fluorescence polarization (FP) assay to measure binding of SRD peptides and Sin3 (figure 1.8C), and found that while mostly other peptides showed binding behavior and dissociation constants similar to the wild-type SID, SRD2 and 5 stood out with increased

affinity – K_d of 12nM and 70nM, respectively. Conveniently, these two peptides were also shown to be good cell penetrators earlier.

Another experiment assessing the interaction between SRD and Sin3 we conducted was affinity pull-down. Here we sub-cloned the Sin3 construct to fuse it to GST separated by His6 and S-tag, cleavable with two proteolytic sites (Figure 1.8A). We then synthesized N-terminally biotinylated versions of active peptides and negative mutants and performed affinity precipitation with the GST fusion construct. Sin3 could indeed be pulled down by SRD2 in a dose-dependent way, agreeing with aforementioned FP results (figure 1.8B). Therefore, we selected SRD2 as the primary peptide on which to implement modifications necessary for native chemical ligation.

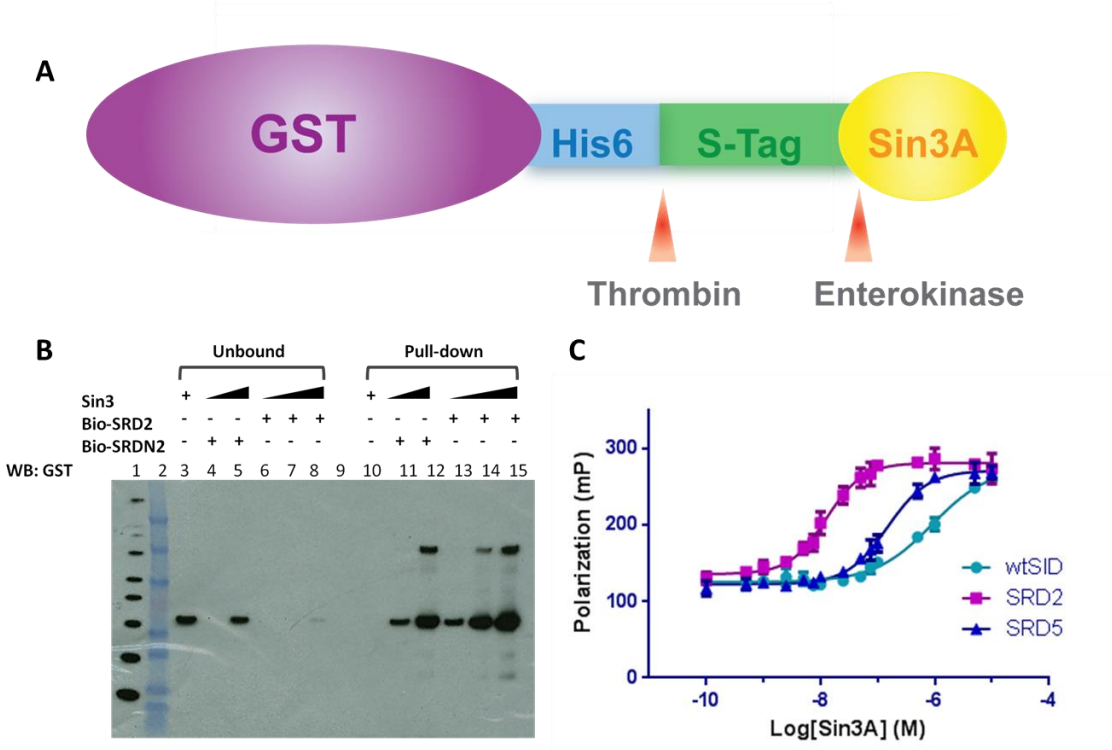


Figure 1.8. Measurements of affinity of SRD peptides to Sin3. (A) Schematic representation of cloning of Sin3A with various affinity tags. (B) Pull-down of Sin3 with biotinylated SRD2 and SRDN immobilized to streptavidin beads. Precipitated Sin3 was visualized via its GST-tag. (C) Fluorescence polarization readings of SRD peptides titrated with Sin3. K_d of wtSID, SRD2, and SRD5 was 900nM, 150nM, and 13nM, respectively.

Chemistry and Biology of Native Chemical Ligation

Traditionally, native chemical ligation (NCL) requires that the N-terminal fragment, in our case the SRD peptide, to have a C-terminal thioester moiety attached to the α-carbon, so that it could react with the N-terminal thiol group on the C-terminal fragment to form a native peptide bond. Here a technical issue arises. Resin for peptide synthesis that can incorporate a C-terminal thioester is necessarily base-labile and thus not compatible with the Fmoc protecting group that is removed by piperidine before every coupling. In the past, researchers who possessed such resources used Boc-based synthesis to make C-terminal thioester peptides[24] (Figure 1.9A). It would be difficult for us to adapt the same method mainly for two reasons: first, such resin requires hydrofluoric acid cleavage, which we could not perform in-house; second, Boc-protected non-natural amino acids for stapling were not available commercially.

As NCL is a field of increasingly active research gaining more and more attention, synthetic routes that eliminate the need for Boc-chemistry and use of HF by varying linker chemistry have been developed[34-37]. We based our selection of new methodology on the criteria of safety, cost, and efficiency, and arrived at a strategy invented by Blanco-Canosa and Dawson in 2008[38,39]. The method utilizes a C-terminal Fmoc-4-diaminobenzoic acid (Fmoc-Dbz) linker on regular rink resin that could be converted to an N-acyl-benzimidazolinone (termed Nbz) leaving group at the end of peptide of peptide synthesis. The Nbz group is capable of undergoing ligation with thiol via an N-acylurea intermediate. Through this route Fmoc-based SPPS is back in business (Figure 1.9B).

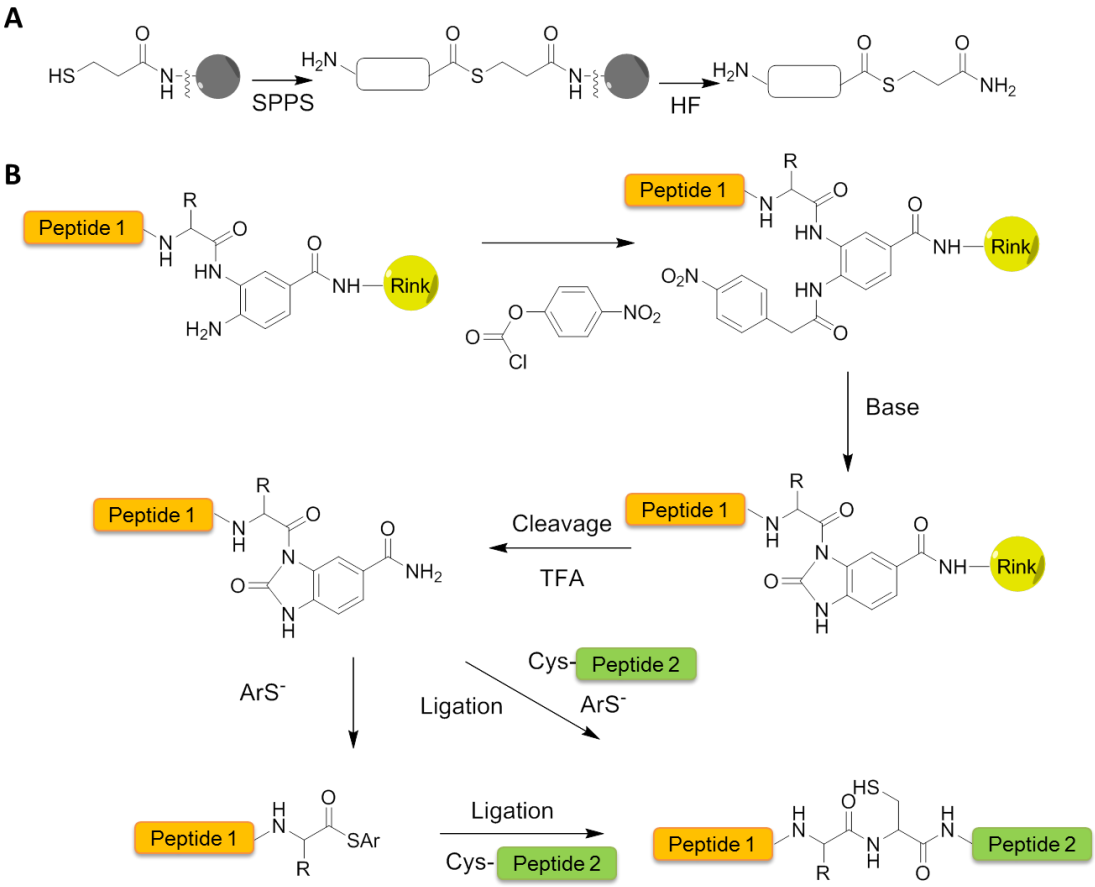


Figure 1.9. Synthetic scheme for C-terminal thioester peptide to use in NCL. (A) Traditional methodology uses a thioester linker between peptide and resin, which necessitates *tert*-butoxycarbonyl solid-phase synthesis (Boc-SPPS) and final cleavage by hydrofluoric acid. (B) New approach developed by Blanco-Canosa and Dawson utilizing an N-terminal acylurea group (-Nbz) and compatible with Fmoc-SPPS.

We first established in our lab the synthesis of the Fmoc-Dbz linker *de novo* as reported, which later became commercially available. We then used the linker to make a model peptide, the wild-type SID, and tried to obtain the final Nbz product, mainly to avoid complications caused by metathesis at this stage. Fortunately, we were able to make both the unmodified and the FITC-labeled version of the wtSID-Nbz after some troubleshooting, as detailed in Experimental Methods.

Almost as foreseen, problems abound when we tried to repeat the procedure while adding the staple. Synthesis of FITC-SRD2-Nbz proved quite challenging, mainly because of incompatibility between metathesis and certain intermediates of the linker. We had to go

through trial-and-error to sample different ordering of reactions to find one particular order that worked: 1) sequence elongation, 2) metathesis 3) activation of linker (-Dbz→-Nbz), and 4) FITC coupling. Finally, SRD2-Nbz and FITC-SRD2-Nbz were made. However, we observed rapid hydrolysis of the Nbz to give a carboxylic acid end during HPLC purification, which was not seen with the wtSID-Nbz peptide, and unfortunately quite detrimental to the yield.

On the C-terminal fragment, i.e. the recombinant protein front, we had in our lab's possession a construct expressing minimized bHLH-Zip of Max, starting with point mutation A12C following the start codon. According to literature[40], such a construct should yield the N-terminal cysteine protein necessary for NCL, as it would undergo cleavage of the first methionine by the endogenous methionine amino-peptidase in *E. coli*. We prepared the protein and tried to perform ligation with wtSID-Nbz and SRD-Nbz, but could not obtain any correct adduct. Clouding the situation further was the lack of accurate analytics to monitor the exact change in molecular weight of molecules: desalting of protein solution was probably not complete (it became especially difficult after ligation when the analyte was in a high salt buffer), resulting in poor ionization and broad peaks in MALDI. Because the MW readout was only exact to about +/- 100Da, we knew for certain that ligation was not happening (which would have resulted in the addition of peptide's MW >2000Da), but could not tell what exactly was happening to the protein. In fact, we could not even be sure if the starting material protein was correct.

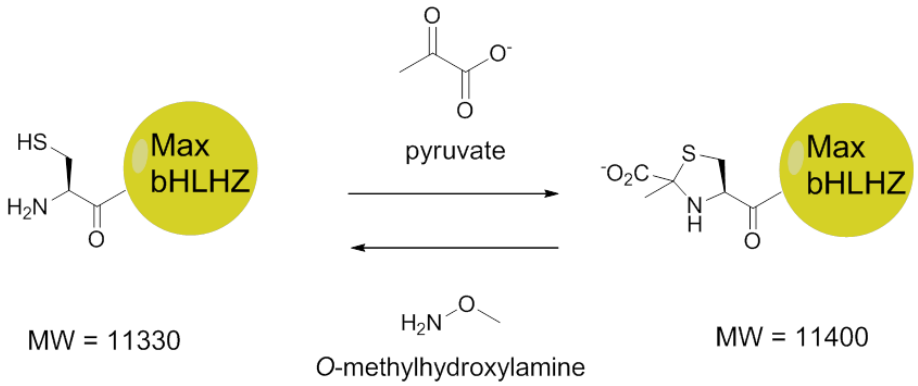


Figure 1.10. Covalent modification of N-terminal cysteine on Max bHLH-Zip. A +70 shift in molecular weight was observed on LC/MS for the starting material Max bHLH-Zip, which matches to a previous reported phenomenon of pyruvate adduct formation on proteins with N-terminal cysteine. The transformation could be reversed by treatment with o-methylhydroxylamine.

In this round of problem-solving, we first designed a control C-terminal fragment: a peptide of sequence CRAFS. As this short synthetic peptide would certainly contain the correct N-terminal free cysteine confirmable by LC/MS, we used it to eliminate the possibility of errors in the N-terminal fragment or ligation conditions. wtSRD-Nbz underwent facile ligation with CRAFS and yielded adduct peptide of the correct MW. Therefore we knew that the source of problem was with bHLH-Zip.

With much effort we found a way of preparation of the recombinant protein that allowed for LC/MS analysis, and the MW we saw explained quite a lot. A +70Da modification was seen, which we now knew to be from addition of pyruvate to the N-cysteine to yield a five-membered ring moiety post-translationally, a little-known but documented phenomenon[41]. After spending time and efforts to optimize experimental conditions, we were able to counteract the unwanted covalent modification by reaction with O-methyl hydroxylamine to regain the free N-Cys protein (figure 1.10).

Now that both components as well as reaction conditions were ready, we finally made our first semi-synthetic protein by NCL by fusing wtSID to bHLH-Zip (figure 1.11). The yield of the reaction was passable though could be improved – about 60-70% of the starting protein

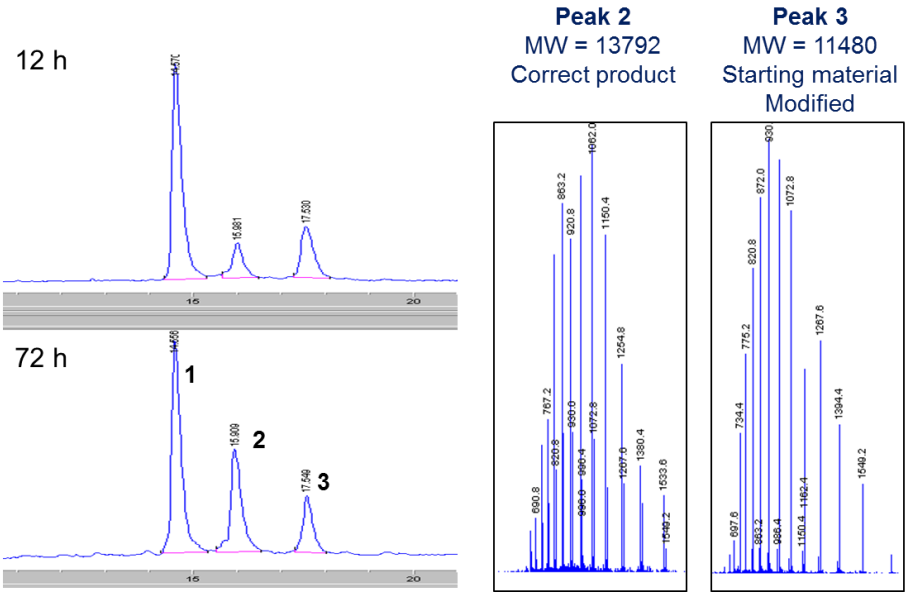


Figure 1.11. Native chemical ligation of wtSID and Max bHLH-Zip. Left, LC/MS trace (MS intensity) of reaction mixture at 12 and 72hr; peak 2 contained the product. Right, m/z ratio of peaks 2 and 3.

was reacted at 72h (peptide-Nbz was in excess). Part of the incompleteness came from trouble brought by the reducing agent. A reducing environment was necessary to keep the N-Cys thiol from oxidation, and a variety of reducing agents were tried including DTT, TCEP, and sodium 2-sulfanylethanesulfonate ("Mesna"). All of them would to some extent form adduct with the ligation intermediate that prevents the rearrangement step leading to the final product (figure1.12). It seemed a necessary evil, though efforts on how to avoid such side reaction would be worthwhile.

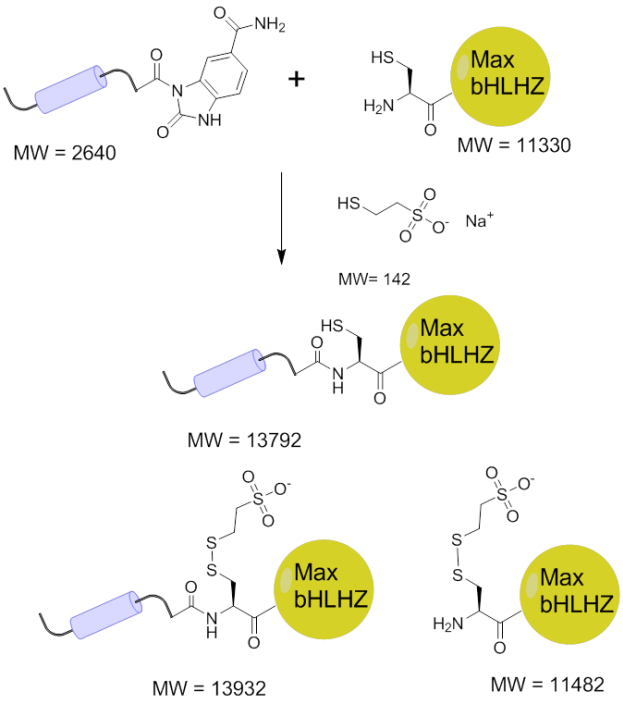


Figure 1.12. Product and by-product formation in NCL. In addition to the correct ligation product, two other entities were formed: ligation intermediate and starting material oxidized by Mesna to disulfide, both of which would stall ligation.

The really coveted product, however, was fusion of SRD2-Nbz and FITC-SRD-Nbz to the Max bHLH-Zip. Unfortunately, using the same reaction conditions established from above, we were not able to obtain substantial amount of this fusion protein. Firstly, the SRD2-Nbz peptide with or without FITC, already scarce to begin with, had very low aqueous solubility. Even at 1mM (concentration required for NCL) in reaction buffer containing 6M guanidinine hydrochloride, the peptide would precipitate after some time.

Secondly, the aforementioned hydrolysis of Nbz happened in the ligation mixture. In summary, SRD2-Nbz was hard to dissolve, and the little amount dissolved would soon be hydrolyzed to acid. The added hydrophobicity from the staple was easier to rationalize then SRD2's much increased susceptibility to hydrolysis in comparison to wtSID. Regrettably this remains a problem that we have not solved to date, and in the meantime other types of ligation chemistry were explored, as will be presented in Chapter II.

Experimental Methods

Peptide Synthesis

Preparation: resin was weighed in reaction vessels, and swollen in NMP for at least an hour. The amount of amino acids used was approximately six times molar excess to the loading of resin for normal amino acids, and three times excess for non-natural amino acids. They were weighed, dissolved and sonicated in appropriate volumes of NMP (sonication might be necessary) to yield 0.4M solution. When not in use, the amino acid solutions were stored at 4°C. At the beginning of each day, fresh solution of PyClock was prepared at concentration of 0.22g/mL in NMP (0.4M). 25% of piperidine in NMP was prepared prior to starting synthesis.

Synthesis: Fmoc was removed with 25% piperidine for 25 minutes. The resin was washed extensively with NMP four times. To drained resin, appropriate volumes of amino acid, PyClock, and DIPEA were added (in this order). Coupling was run for 45 minutes for regular amino acids and at least one hour for beta-branched amino acids (V, P, I, T), H, and all non-natural amino acids. At the end of the coupling, resin was drained and washed thoroughly in NMP.

Metathesis: resin was washed first with NMP, then DCM, finally DCE. Grubb's Catalyst I was warmed to room temp. Needed amount (~7% of reaction scale) was weighed and dissolved in DCE and added to the resin. Reaction was run for 1.5 to 2 hours. The process was repeated once with freshly prepared catalyst.

Circular Dichroism Measurement of Peptide Helicity

The extent of helix induction in stapled peptides was measured in phosphate-buffered saline solution (pH 7.4) by circular dichroism spectropolarimetry, using the characteristic

minima at 208 and 222 nm as a well-established measure of helical character. Less commonly there would be stapled peptides exhibiting other structural traits, for instance, 310-helix characterized by a minimum centered at 207nm and a negative shoulder between 220 and 230 nm. The CD readout was converted to molar ellipticity by inputting the concentration of peptide in solution, usually around 50 μM. When solubility of a peptide was too low to be dissolved in PBS, milli-Q water or a combination of water and acetonitrile was used. Besides secondary structure, CD experiments were also used in other instances to determine the thermal stability and melting point of a peptide via continuous recording of the spectrum over rising temperature.

Measure of Peptide Cellular Uptake

Hela cells cultured in complete growth medium consisting of DMEM, 10% FBS with penicillin/streptomycin were plated on chamber slides and grown to greater than 80% confluence and medium was removed. FITC-labeled Peptides were dissolved at appropriate concentrations in complete growth medium and added to cells. A DMSO control was also prepared following the highest concentration of peptide vehicle. Cells were then incubated at 37 °C for desired length of time, at the end of which medium was aspirated and cells washed extensively in cold PBS. Cells were fixed using 4% formaldehyde and washed again in PBS. The cover slips were mounted using VectaShield Hardset with DAPI. FITC and DAPI fluorescence was monitored by confocal microscopy.

Cloning, Expression, and Purification of Sin3

The original Sin3 PAH2 construct in a pET24a vector was kindly gifted to us by Prof. Radhakrishnan's group at Northwestern University. This construct did not contain any

affinity tag. The protein was subsequently cloned into a pET-41 Ek/LIC Vector using the Novagen ligation-independent cloning (LIC) kit. Oligos used in the cloning PCR were

Sin3_LIC_Fwd, 5' GACGACGACAAGATGTCTCTGCAAAACAATCAGCCT 3' and Sin3_LIC_Rev,

5' GAGGAGAAGCCCGGTTAGGCATCTGGCAGGAACTGTCC 3'. The new construct was fused to N-terminal GST, His6, and S-tag.

The resulting expression plasmid was transformed into Rosetta 2 (DE3) pLysS chemically competent cells using kanamycin and chloramphenicol selection. A single colony was picked to start an overnight LB culture at 37°C. 10mL of starter culture was used to inoculate to flasks containing 1L of LB. Culture was grown at 37°C for approximately 4 hours until an OD_{600} of 0.6-0.8 was reached. The temperature was decreased to 30°C and expression was induced by the addition of IPTG to a final concentration of 0.1mM. After another 4-6 hr of growth at 30°C, cells were harvested by centrifugation, flash frozen by liquid nitrogen, and stored at -80°C until use.

Frozen cell pellet from 1L culture was thawed on ice, and ice cold Lysis Buffer was added to pellet, 25-30mL/1L culture. Cells were resuspended by gentle pipetting and sonicated to lyse open: 10 sec on, 30 sec off for 2 min, wait for 1 minute, then repeated once. Lysed cells were entrifuge at 14000 rpm for 40 min. During centrifugation, Ni-NTA resin was equilibrated with lysis buffer (2-3mL slurry per 1L culture). Lysate was filtered with 2um filter, transfered to a 50mL conical tube or gravity-pull column, add equilibrated resin, and incubated on rock bed in cold room for 1h, turning every 15 min. The flowthrough was drained, and the resin was washed once with 25mL of Wash Buffer. The bound proteins were eluted by washing the resin once with 1.5mL of Elution Buffer followed by 5 elutions with 2.5mL of Elution Buffer.

The elution was further purified by Superdex 200 Gel Filtration. Fractions were analyzed by SDS-PAGE and combined to yield GST-Sin3 in >90% purity.

Lysis Buffer
20mM HEPES pH 7.5
150mM KCl
1mM EDTA
5% Glycerol
1mM PMSF

Wash Buffer
Lysis Buffer, 50mM imidazole

Elution Buffer
Lysis Buffer, 300mM imidazole

Gel Filtration Buffer
20mM HEPES pH 7.5
150mM NaCl
5% Glycerol

Fluorescence Polarization

The FP Assay had to be performed within a week of purifying the protein or thawing of flash frozen protein aliquots. Sin3 protein sample in Gel Filtration Buffer had been quantified prior to the assay using the standard BCA assay and double-checked by absorbance reading at 280nm, then diluted in the same buffer to 3μM working stock. The assay was performed using a standard opaque 384-well plate in which serial dilution of protein was made from the 3μM starting point and diluting at the interval of 3X, until the last protein sample reached a final concentration of 0.1nM. Peptides were also dissolved in Gel Filtration Buffer and added to each well of protein at the final concentration of 15nM and total final volume of <90μL. Each data point was done in triplicate. The plate was sealed with aluminum foil and incubated in the cold room for 0.5-1h.

At the end of the incubation period, fluorescence polarization values were obtained from a Hewlett-Packard SpectraMax plate-reader using the "Endpoint"→ "FP" setting and excitation/emission wavelengths of 485nm and 525nm.

Pull-down Assay

Pull-down assays were also done with protein samples that were no longer than a week old from purification or thawing.

Preparation: Life Technology M280 magnetic streptavidin-coated beads were equilibrated in Binding Buffer and drained using the magnetic strip. Biotinylated peptides dissolved in Binding Buffer at 10μM were added to the beads, along with one sample treated with DMSO control in Binding Buffer. Immobilization of peptides was done by gentle nutating overnight in the cold room.

Pull-down: on the next day, the beads were drained and washed thoroughly in Wash Buffer. Sin3 protein sample was diluted to desired concentration in Binding Buffer and introduced to the peptide-bound or mock beads. Pull-down was allowed to go for up to 3h by gentle nutating in the cold room. All samples were then drained and washed extensively in Wash Buffer. At the end of the last wash, 1X SDS Loading Buffer was added to the beads and the mixture was boiled at 95°C for at least 5 minutes to release bound protein. Samples were then run on SDS-PAGE gel to be visualized by Western blot.

Immunoblotting: At the completion of gel electrophoresis, SDS-PAGE was transferred to nitrocellulose membrane using the wet transfer method in Transfer Buffer. After transfer, the blot was blocked in Blocking Buffer for up to 1h at room temperature, and incubated with primary antibody (rabbit GST Antibody #2622, Cell Signaling) at 1:1000 in Blocking Buffer overnight in the cold room. On the next day, the blot was washed extensively and

then incubated with secondary antibody (Anti-rabbit IgG, HRP-linked Antibody #7074, Cell Signaling) in Blocking Buffer for 1h at room temperature. The blot was again washed thoroughly before visualization by addition of SuperSignal West Pico Chemiluminescent Substrate from Pierce.

Binding Buffer
Phosphate Buffered Saline pH 7.4 (NaCl = 147mM, regular)
0.01%(v/v) Tween-20
0.5%(w/v) BSA

Wash Buffer
Phosphate Buffered Saline pH 7.4 (NaCl = 500mM, high salt)
0.01%(v/v) Tween-20

SDS Loading Dye, 4X
50 mM Tris-HCl pH 6.8
2% SDS
10% glycerol
1% β-mercaptoethanol
12.5 mM EDTA
0.02 % bromophenol blue

Transfer Buffer, 10X, for 4 L
121.1 g Tris base
576 g glycine
Bring up the volume to 4 L with ddH_2O
(Add 20% methanol when making 1X solution)

Blocking Buffer
Tris Buffered Saline pH 7.4
0.1%(v/v) Tween-20
5% (w/v) BSA

Synthesis of –Nbz Peptides

Preparation and coupling of first residue: the Dawson Dbz AM resin (EMD Chemicals) was pre-swollen in NMP. Fmoc was removed by 20% piperidine in NMP. To couple the first residue, amino acid and HOBt and HATU (6 equivalents of resin loading) and DIPEA (9 equivalents of resin loading) were added in that order and coupling was allowed to proceed for 1h.

Synthesis and activation: the elongation of sequence was done in regular Fmoc-SPPS manner as described earlier. Importantly, ring-closing metathesis needed to take place before the activation of linker when the free amine was intact. Upon coupling of the second-to-last residue, the N-terminal residue was introduced using a Boc-amino acid. For activation of linker, p-nitrophenyl chloroformate (5 equivalents of resin loading) was added in DCM and left to gently agitate under nitrogen for 1h. Resin was drained and washed and 0.5M DIPEA was added in 10mL DMF and left to react for 30min. After installation of the –Nbz group, FITC could be coupled to the N-terminus of the peptide. In general, the order of installations should be as follows: 1) sequence elongation, 2) metathesis 3) activation of linker (-Dbz→-Nbz), and 4) FITC coupling. The finished peptide was then cleaved in the same TFA cocktail and purified on HPLC as mentioned before.

Expression and Purification of Max bHLH-Zip

A pET28a plasmid containing the minimized Max bHLH-Zip with N-terminal cysteine (obtained from Dr. Eileen Kennedy and Dr. Jerry Hilinski in the lab) was transformed into Rosetta 2 (DE3) pLysS chemically competent cells using kanamycin and chloramphenicol selection. A single colony was picked to start an overnight LB culture at 37°C. 10mL of starter culture was used to inoculate to flasks containing 1L of LB. Culture was grown at 37°C for approximately 4 hours until an OD_{600} of 0.6-0.8 was reached. The temperature was decreased to 18°C and expression was induced by the addition of IPTG to a final concentration of 0.1mM. After another 12 hr of growth at 18°C, cells were harvested by centrifugation, flash frozen by liquid nitrogen, and stored at -80°C until use.

Frozen cell pellet from 1L culture was thawed on ice, and ice cold Lysis Buffer was added to pellet, 25-30mL/1L culture. Cells were resuspended by gentle pipetting and sonicated to lyse open: 10 sec on, 30 sec off for 2 min, wait for 1 minute, then repeated once. Lysed cells

were entrifuge at 14000 rpm for 40 min. During centrifugation, Ni-NTA resin was equilibrated with lysis buffer (2-3mL slurry per 1L culture). Lysate was filtered with 2um filter, transfered to a 50mL conical tube or gravity-pull column, add equilibrated resin, and incubated on rock bed in cold room for 1h, turning every 15 min. The flowthrough was drained, and the resin was washed twice with 45mL of Wash Buffer. The bound proteins were eluted by washing the resin twice with 2.5mL of Elution Buffer. (Buffer compositions were the same as used in purification of Sin3).

Native Chemical Ligation

To prepare N-cysteine Max for native chemical ligation, acetone was added to the purified protein in Elution Buffer at 4:1 to precipitate the protein. The pellet was dissolved in phosphate buffered saline pH 7.4 containing 400mM o-methylhydroxylamine to convert the cyclized pyruvate-adduct back to the free thiol form. The corrected protein was again precipitated by acetone and re-dissolved in Ligation Buffer. The final concentration of Max in Ligation Buffer was about 0.2mM.

wtSID-Nbz from 10mM stock in DMSO was added to the Max-containing Ligation Buffer at final concentration of 10mM. The mixture was stirred at room temperature for 12 – 72h and monitored by LC/MS.

<u>Ligation Buffer</u>
300mM $NaHPO_4$ pH 7.8
6M guanidinium hydrochloride
100mM Mesna

References

1. Daley, G.Q. Gleevec resistance: lessons for target-directed drug development. *Cell Cycle* **2**, 190-1 (2003).
2. Hopkins, A.L. & Groom, C.R. The druggable genome. *Nat Rev Drug Discov* **1**, 727-30 (2002).
3. Schreiber, S.L. Organic chemistry: Molecular diversity by design. *Nature* **457**, 153-4 (2009).
4. Henchey, L.K., Jochim, A.L. & Arora, P.S. Contemporary strategies for the stabilization of peptides in the alpha-helical conformation. *Curr Opin Chem Biol* **12**, 692-7 (2008).
5. Cheng, R.P., Gellman, S.H. & DeGrado, W.F. beta-Peptides: from structure to function. *Chem Rev* **101**, 3219-32 (2001).
6. Liu, J., Wang, D., Zheng, Q., Lu, M. & Arora, P.S. Atomic structure of a short alpha-helix stabilized by a main chain hydrogen-bond surrogate. *Journal of the American Chemical Society* **130**, 4334-7 (2008).
7. Smith, B.A. et al. Minimally cationic cell-permeable miniature proteins via alpha-helical arginine display. *Journal of the American Chemical Society* **130**, 2948-9 (2008).
8. Schafmeister, C.E., Po, J. & Verdine, G.L. An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides. *Journal of the American Chemical Society* **122**, 5891-5892 (2000).
9. Walensky, L.D. et al. Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix. *Science* **305**, 1466-70 (2004).
10. Verdine, G.L. Drugging the "undruggable". *Harvey Lect* **102**, 1-15 (2006).
11. Moellering, R.E. et al. Direct inhibition of the NOTCH transcription factor complex. *Nature* **462**, 182-8 (2009).
12. Verdine, G.L. & Hilinski, G.J. Stapled peptides for intracellular drug targets. *Methods Enzymol* **503**, 3-33 (2012).
13. Grossmann, T.N. et al. Inhibition of oncogenic Wnt signaling through direct targeting of beta-catenin. *Proc Natl Acad Sci U S A* **109**, 17942-7 (2012).
14. Henriksson, M. & Luscher, B. Proteins of the Myc network: essential regulators of cell growth and differentiation. *Adv Cancer Res* **68**, 109-82 (1996).
15. Grandori, C., Cowley, S.M., James, L.P. & Eisenman, R.N. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Biol* **16**, 653-99 (2000).
16. Massari, M.E. & Murre, C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* **20**, 429-40 (2000).
17. Ayer, D.E., Lawrence, Q.A. & Eisenman, R.N. Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* **80**, 767-76 (1995).
18. James, L. & Eisenman, R.N. Myc and Mad bHLHZ domains possess identical DNA-binding specificities but only partially overlapping functions in vivo. *Proc Natl Acad Sci U S A* **99**, 10429-34 (2002).
19. McMahon, S.B., Van Buskirk, H.A., Dugan, K.A., Copeland, T.D. & Cole, M.D. The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* **94**, 363-74 (1998).
20. Soucek, L. et al. Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis. *Cancer Res* **62**, 3507-10 (2002).
21. Soucek, L. et al. Modelling Myc inhibition as a cancer therapy. *Nature* **455**, 679-83 (2008).

22. Nair, S.K. & Burley, S.K. X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* **112**, 193-205 (2003).
23. Sauve, S., Tremblay, L. & Lavigne, P. The NMR solution structure of a mutant of the Max b/HLH/LZ free of DNA: insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. *J Mol Biol* **342**, 813-32 (2004).
24. Hackenberger, C.P.R. & Schwarzer, D. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angewandte Chemie-International Edition* **47**, 10030-10074 (2008).
25. Dawson, P.E., Muir, T.W., Clark-Lewis, I. & Kent, S.B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-9 (1994).
26. Holford, M. & Muir, T.W. Adding 'splice' to protein engineering. *Structure* **6**, 951-6 (1998).
27. Severinov, K. & Muir, T.W. Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J Biol Chem* **273**, 16205-9 (1998).
28. Muir, T.W., Sondhi, D. & Cole, P.A. Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* **95**, 6705-10 (1998).
29. Camarero, J.A., Cotton, G.J., Adeva, A. & Muir, T.W. Chemical ligation of unprotected peptides directly from a solid support. *J Pept Res* **51**, 303-16 (1998).
30. Macmillan, D. & Arham, L. Cyanogen bromide cleavage generates fragments suitable for expressed protein and glycoprotein ligation. *Journal of the American Chemical Society* **126**, 9530-1 (2004).
31. Erlanson, D.A., Chytil, M. & Verdine, G.L. The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem Biol* **3**, 981-91 (1996).
32. Brubaker, K. et al. Solution structure of the interacting domains of the Mad-Sin3 complex: implications for recruitment of a chromatin-modifying complex. *Cell* **103**, 655-65 (2000).
33. Jullian, M. et al. N-terminus FITC labeling of peptides on solid support: the truth behind the spacer. *Tetrahedron Letters* **50**, 260-263 (2009).
34. Shin, Y. et al. Fmoc-based synthesis of peptide-(alpha)thioesters: Application to the total chemical synthesis of a glycoprotein by native chemical ligation. *Journal of the American Chemical Society* **121**, 11684-11689 (1999).
35. Mezo, A.R., Cheng, R.P. & Imperiali, B. Oligomerization of uniquely folded mini-protein motifs: Development of a homotrimeric beta beta alpha peptide. *Journal of the American Chemical Society* **123**, 3885-3891 (2001).
36. Warren, J.D., Miller, J.S., Keding, S.J. & Danishefsky, S.J. Toward fully synthetic glycoproteins by ultimately convergent routes: A solution to a long-standing problem. *Journal of the American Chemical Society* **126**, 6576-6578 (2004).
37. Zeng, Y., Ramya, T.N.C., Dirksen, A., Dawson, P.E. & Paulson, J.C. An Efficient and Quantitative Chemical Method for Tagging Sialylated Glycoproteins on Living Cells through Aniline Catalyzed Oxime Ligation. *Glycobiology* **18**, 998-999 (2008).
38. Blanco-Canosa, J., Ingale, S. & Dawson, P.E. Native Ligation with Post-Translational Modifications Using C-Terminal, Nacylurea Peptides. *Biopolymers* **92**, 371-371 (2009).
39. Blanco-Canosa, J.B. & Dawson, P.E. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angewandte Chemie-International Edition* **47**, 6851-6855 (2008).

40. Hirel, P.H., Schmitter, M.J., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A* **86**, 8247-51 (1989).
41. Gentle, I.E., De Souza, D.P. & Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug Chem* **15**, 658-63 (2004).

Chapter II – Experimental Optimization of Semi-synthetic Transcriptional Factors

Adapting a New Bio-conjugation Method

Although native chemical ligation is an attractive way to create fusion proteins that does not introduce any non-peptide element in the product, it is neither the only method to connect a synthetic peptide to a recombinant protein nor the most efficient. A litany of chemoselective ligation or conjugation strategies have been developed in recent decades[1-3], most of which utilize various types of naturally occurring nucleophiles from one building block, such as the amine of lysine or thiol of cysteine, to connect to artificially installed electrophiles on the other.

Quite recently one of these methods, the maleimide – cysteine conjugation[4-8], was successfully used to bring together two different stapled peptides in another target system. In the setup, depicted by figure 2.1, one peptide contains a lysine residue that was orthogonally protected by an Mmt group that can be removed by dilute trifluoroacetic acid. The free amine of the lysine side chain can then be labeled with Succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) to yield a maleimide-activated peptide, leaving the N-terminus free for other use, e.g. fluorophore labeling. The purified

maleimide peptide reacts with the thiol group of the other, cysteine-containing peptide. Although a non-native moiety is left in the fusion product, a rather undesirable trait, this method offers a few advantages over NCL: the building blocks are easier to prepare and the reaction is more facile and foolproof. Furthermore, it allows the cysteine to be anywhere in the C-terminal fragment rather than confining it to the N-terminus, which greatly would simplify things considering how labile and promiscuous an N-Cys could be as we witnessed earlier.

Encouraged by the successful use of this cysteine-maleimide conjugation method in linking two synthetic peptides, we proceeded to adapt it for making SRD2-bHLH-Zip. FITC-SRD2 was re-synthesized to include an Mmt-protected lysine on the C-terminus, which upon sequence completion and N-terminal FITC-labeling was removed to install the maleimide

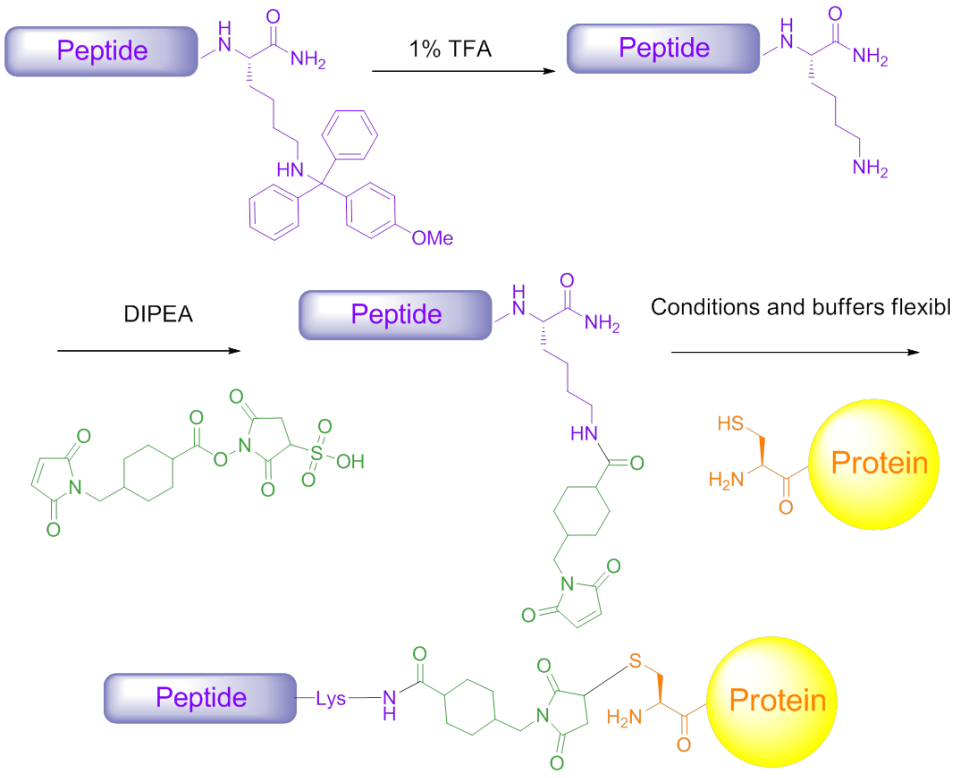


Figure 2.1. Proposed synthetic scheme of cysteine-maleimide conjugation. The N-terminal peptide is functionalized with maleimide on a lysine side chain. The C-terminal recombinant protein, prepared via the same way as in native chemical ligation, contains an N-terminal cysteine. Conjugation is achieved by addition of sulfhydryl to maleimide.

group. On the protein end, Max bHLH-Zip was prepared much the same way as before. For the choice of reaction buffer, we found that although FITC-SRD2-maleimide was slightly less hydrophobic that the –Nbz version of the peptide, it was still not compatible with most commonly used buffers for proteins such as Tris- or phosphate-buffered saline. Fortunately it did not precipitate at 1mM in 6M guanidinium hydrochloride at least. As a compromise, however, the recombinant protein would have to be denatured for conjugation.

After screening for conditions, we were then able to link FITC-SRD2-maleimide to N-cys Max bHLH-Zip. Remarkably, correction of the covalently modified N-cysteine and subsequent conjugation could take place in one single step without isolation of the intermediate free thiol-containing protein (figure 2.2). The "one-pot" reaction worked quite well – almost all of the starting protein would form the correctly sized conjugate with the peptide within 12 hours as monitored by the LC/MS (Figure 2.3). The product was stable and could be easily purified on HPLC and lyophilized to be obtained in powder form. We soon found that the reaction was also quite scalable, and before long over 10mg of semi-synthetic protein was produced. Being able to make the chemistry work in such a clean and efficient way was very exciting, and immediately we wanted to immediately test the biological effects – cell permeability now that we had a stapled, FITC-labeled compound available. However, at this point we ran into obstacles again. Solubility, it

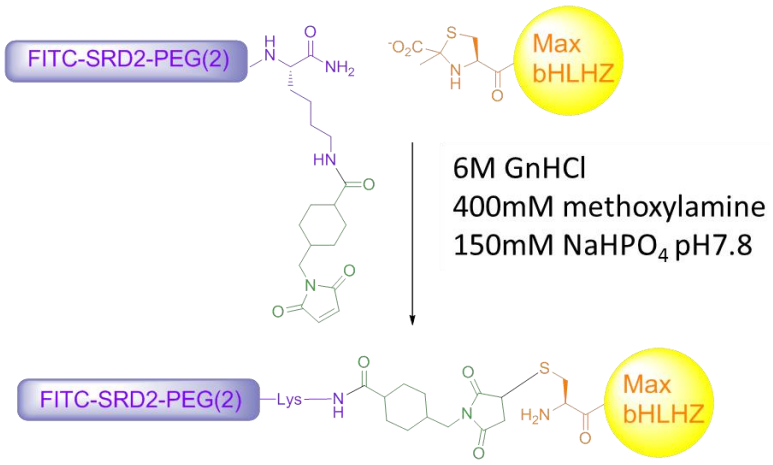


Figure 2.2. "One-pot" conjugation of SRD and Max. Optimal conditions allowed elimination of the extra step to reverse cyclization of N-terminal cysteine; Purified Max containing the pyruvate adduct could be used directly for conjugation

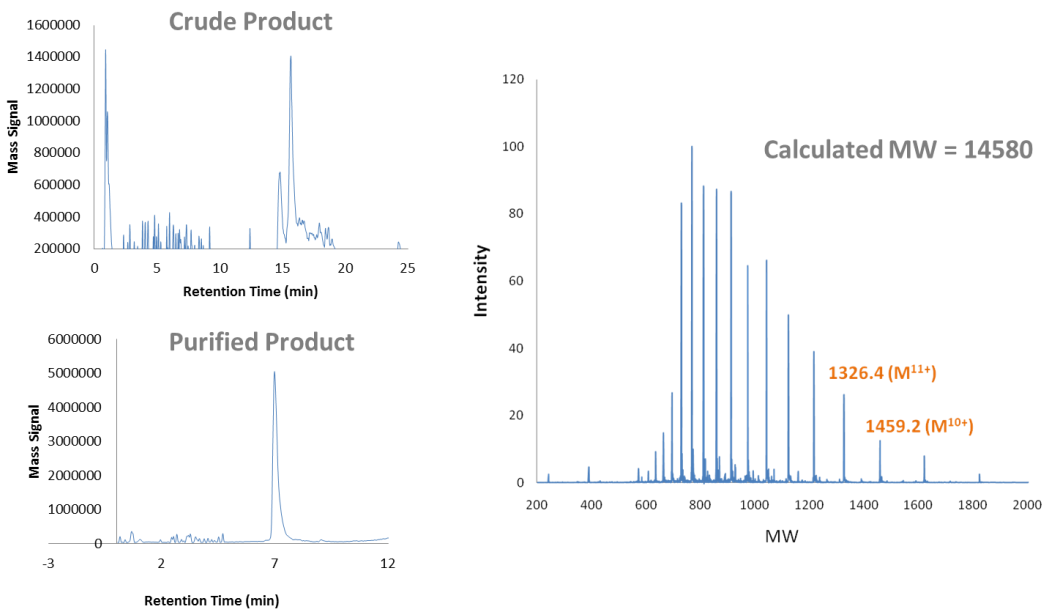


Figure 2.3. LC/MS analysis of conjugation and subsequent purification. Left, LC/MS trace (MS intensity) of reaction mixture after 12h showed robust conjugation. Product was purified by HPLC. Right, m/z ratio of product peak.

seemed, was an inescapable issue with this system. The conjugate could be dissolved only in pure deionized water (Milli-Q, pH~6.0) or DMSO. When the stock solution was introduced to any other aqueous buffer within the physiological pH range or even tissue culture medium, it would precipitate immediately. And once the solution was centrifuged to remove the precipitate from the supernatant, the latter had undetectable absorbance at 494nm (FITC) and hence no detectable amount of conjugate. Incidentally, acidifying the buffer to below pH 6 also helped the protein dissolve, although at that pH no measurement would be relevant.

Many attempts such as sonication, heating the solution up, or adding in trace amounts of stock solution sequentially were made but none were effective at solubilizing the conjugate. Since it was likely that the conjugate, being formed under denaturing conditions and worked up in organic solvents, never had a proper chance to refold which was at least part of the reason for precipitation. Slightly more sophisticated experiments were done in an effort to promote folding, such applying the conjugate first in Milli-Q water to Ni-NTA beads

or size-exclusion chromatography column and eluting slowly by slowing exchanging to a more suitable buffer. These attempts did not work out either, since the conjugate would seemingly aggregate non-specifically on the resin and not elute with these buffers.

At this point, it was clear to us that the unmanageable physical properties of the conjugate were not a trivial problem and there were no obvious solutions. We took a hypothesis-based approach and dissected the troubleshooting to two parts, looking at each fragment of the conjugate separately, making informed guesses on where the biggest driver to insolubility was, and trying to engineer modifications in that element that lead to more desirable properties.

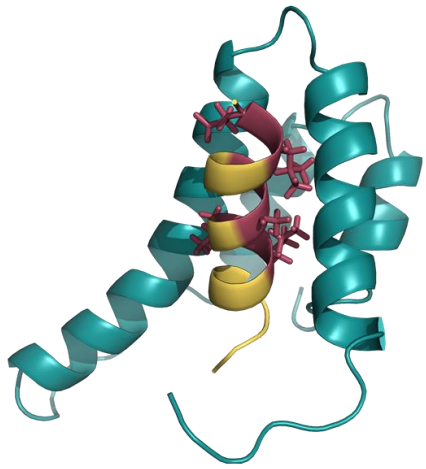
Improving the solubility of Stapled Repression Domain peptides

Our first hypothesis was that the low solubility of SRD2 and all other SRD peptides we have made so far as their own entities had been carried over to the conjugate. Even if SRD2 represents only a small part in the conjugated product, it could still significantly decrease the overall solubility in a number of possible ways: forming micelles, promoting aggregation, preventing the rest of the semi-synthetic protein from folding properly, etc. Thus we decided that re-designing and extending the peptide library to obtain more soluble peptides would be likely to result in substantial improvements in conjugate solubility while incurring relatively small costs in time and resources.

As no precise method for predicting the solubility of a peptide using just the primary sequence (or even crystallographic/solution NMR structure) as the input existed to our knowledge, we made the simplifying assumption that the overall hydrophobicity would be positively correlated with the number of hydrophobic residues based on all the evidence available. This assumption did not reflect whether the exact relationship was a linear

addition, multiplication, or a more complex function with other variables at play, which is the most likely since given the helical shape of the peptide, the spatial coordinates of these hydrophobic residues and their interaction with neighbors would come in to play as well. If we had had that information and thus a prediction method that could differentiate the net contribution to overall insolubility by each hydrophobic residue, we would have known which ones to focus on to change into something less hydrophobic without sacrificing too much affinity to the target protein. Since we did not, we weighted the hydrophobic residues equally and focused our efforts on those that appeared to be less directly involved in Van der Waals interactions for binding (figure 2.3A). Revisit of the solution structure showed

A



C

Compound	Abs	Solubility (μM)
SRD2	0.032	0.42
SRD11	0.115	1.49
SRD12	0.128	1.66
SRD5	0.322	4.18
SRD13	0.128	1.66
SRD14	0.725	9.42

B

Compound	Sequence																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
wtSID	V	R	M	N	I	Q	M	L	L	E	A	A	D	Y	L	E	R	R	E	R	E	A	E	H
Min. Motif		X	X	X	Φ	Z	Z	Φ	Φ	X	A	A	X	X	Φ	E	X							
SRD2	βAla- V	R	R	R	I	*	M	L	L	*	A	A	N	Y	L	E	R							
SRD11	βAla-		R	R	I	*	M	L	L	*	A	A	N	Y	L	E	R							
SRD12	βAla-		R	R	I	*	M	L	L	*	A	A	N	R	L	E	R							
SRD5	βAla- V	R	R	R	I	Q	M	L	L	E	A	A	N	Y	*	E	R	R	*	R				
SRD13	βAla-		R	R	I	Q	M	L	L	E	A	A	N	Y	*	E	R	R	*	R				
SRD14	βAla-		R	R	I	Q	M	L	L	E	A	A	N	R	*	E	R	R	*	R				

Figure 2.4. Second-generation SRD peptides aimed to increase solubility. (A) Reviewing of the crystal structure of SID bound to Sin3 showed that most hydrophobic residues in SID were important for binding, and the few seemingly expendable ones were near N- and C-termini. (B) Sequences of new SRD peptides derived from previous hits, SRD2 and SRD5. (C) Aqueous solubility of SRD was measured by saturating the FITC-labeled peptides in PBS and reading the absorbance of supernatant.

that the short SID helix contained many well-defined hydrophobic interactions with Sin3. We deemed two residues to be relatively expendable: valine at the N-terminus and tyrosine on the C-terminal part of the helix. The valine did not appear to be involved in much of any interactions, while the tyrosine looked slightly riskier to remove. Based on this thinking we made SRD11-14 (figure 2.3B) based on the sequences of SRD2 and SRD5 (best binders from initial library) by removing the N-terminal valine or mutatint the tyrosine into a charged, hydrophilic residue (arginine).

We measured the solubility of SRD11-14 in phosphate-buffered saline, and performed the same fluorescence polarization assay as before to monitor change in affinity to Sin3. We found that in comparison to their parent peptides SRD2 and SRD5, the new SRDs in general were more soluble. SRD11 and SRD12, based on the extremely insoluble SRD2, saw a mild improvement on solubility. SRD 13 and SRD14 fared better in part because their base, SRD5, was less insoluble. We did not observe much adverse effect on affinity (figure 2.5) brought by the mutations. In fact, SRD13 and SRD14 bind with a lower K_d than SRD5. We found this surprising especially in the case of SRD13 vs. SRD5, the only difference being SRD13 had one less amino acid (N-terminal valine). It was possible that the shorter helix became more structured after stapling than the original one, and the gain in secondary structure translated to higher affinity.

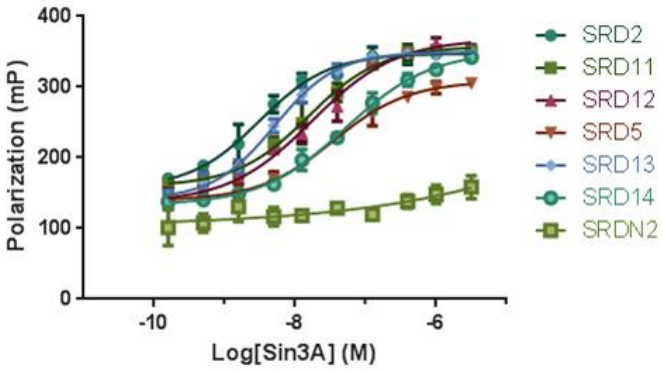


Figure 2.5. Measurement of binding affinity of new SRD peptides to Sin3. Fluorescence polarization readings of SRD peptides titrated with increasing concentration of Sin3. Second-generation SRD peptides retained affinity to Sin3 albeit more soluble than their parent compounds.

Continuing on testing the earlier hypothesis, we constructed

maleimide-labeled derivatives of SRD11-14 and conjugated these to Max bHLH-Zip as before. The result was that only the conjugate containing the most soluble peptide, SRD14-Max, could be diluted into culture medium from water in the presence of albumin without visible precipitation at 10μM. Interestingly this observed solubility in culture medium was higher than that in phosphate buffered saline, perhaps due to stabilizing effects of co-solutes such as albumin. However, when we made the attempt to test cellular uptake, even SRD14 could not stay in solution at 10μM long enough for us to observe intracellular fluorescence – it would precipitate on the cell membrane that prevented further monitoring of intracellular fluorescence. On this end, the amount of change that was imposed on the peptide did increase solubility of the conjugate slightly, but not enough to overcome the hurdle.

Troubleshooting in Preparation of Max bHLH-Zip

Preventing Aggregation of Max Our other hypothesis was that denaturation of Max by guanidine was also a likely culprit of aggregation and insolubility. Therefore the direction was to modify the methodology of the preparation of protein Max bHLH-Zip such it was maintained under native conditions at all time, including conjugation. Firstly, in order to avoid the N-terminal cysteine modification that inevitably required a resolution involving denaturants, a mutant of Max was made to insert an alanine at the N-terminus before cysteine ("ACMax"). During the purification process we noticed something unusual. After Ni-NTA affinity chromatography, which had been sufficient to yield very pure protein in the

past, the gel indicated that another round of purification was necessary. We proceeded with size exclusion chromatography and saw that ACMax had come out in the dead volume, i.e. it was aggregated. This news was of great significance because we had not expected the "native" Max bHLH-Zip to be prone to aggregation, and could be another reason why the conjugate was prone to precipitation. Through our investigation it was found that ACMax was co-eluted from Ni-NTA column with massive amounts of DNA, even though DNase I had been added to the lysis buffer. Being a DNA-binding protein, ACMax's strong affinity to DNA was reasonable but still quite problematic. At length, extra purification by heparin column was also able to separate the protein from DNA, and ACMax no longer eluted as an aggregated mass on size exclusion column (figure 2.6).

On the peptide front, only SRD14 is soluble in the protein buffer in the μM range. As the reaction mixture developed turbidity, it was centrifuged and both the supernatant and precipitate (needed 10% TFA to dissolve in acetonitrile/water) were analyzed by LC/MS. The supernatant was void of any compound, and in the precipitate there were Max dimer and

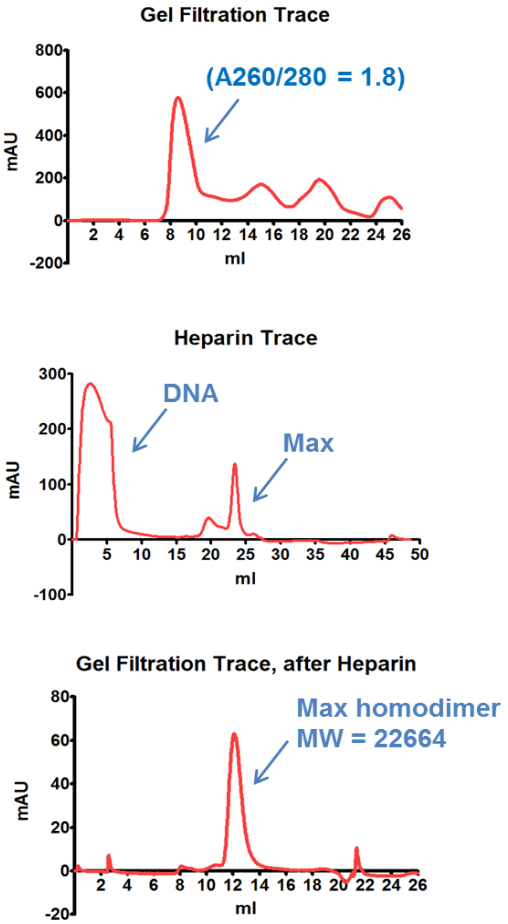


Figure 2.6. Removal of DNA from recombinant Max and de-aggregation. Top, gel filtration of Max after His-tag purification showed aggregation of protein bound to DNA. Middle, DNA could be separated from Max using a heparin column. Bottom, after heparin column Max was no longer aggregated.

peptide-TCEP adduct – it appeared that upon depletion of the reducing agent by maleimide-peptide, Max formed disulfide crosslink. In a separate effort, Omomyc, a homodimerizing variant of Myc[3] was expressed as a GST fusion as an alternative bHLH-Zip to Max. It was designed such that upon TEV cleavage and removal of GST, a cysteine would be revealed at the N-terminus of Omomyc. Omomyc quickly precipitates after cleavage, however – not surprising given the poor stability of wild-type Myc on its own.

Conjugation of Max to Other Peptides We selected from our depository of compounds the top cell-permeable peptides (CPP) as quantified by high-throughput epifluorescence screening (Verdine Lab, unpublished), functionalized them with maleimide and conjugated with Max. Of these CPP-Max conjugates, the SAH-P53-8 (high affinity binder to hDM2)[9] version is particularly interesting since it would conceivably effect targeted ubiquitination of the Myc-Max heterodimer. To date, though all conjugates have been synthesized and purified, insolubility of the lyophilized product, which we had at first attributed partly to the hydrophobicity of SRD, seems a universal trait. It may be hypothesized that Max conjugates do not spontaneously refold when rehydrated (unlike small, simple stapled-peptides) but instead aggregate and precipitate.

Attempts at Refolding SRD14-Max

Because both of the previous two hypotheses could not be tested fully due to insufficient increase in SRD solubility, we would have to continue hunting for more variations in the peptide that might lead to higher gains in solubility, the uncertainty of which had been too great to justify further consumption of time and capital. On the other hand, through the exercise of implementing changes on the peptide, protein, and conjugation conditions, we have made observations that seemed to point collectively at one metric that was particularly non-susceptible to change: the denatured state of the final product. Regardless of what was causing the denaturation, aggregation, and proclivity to precipitate, the end

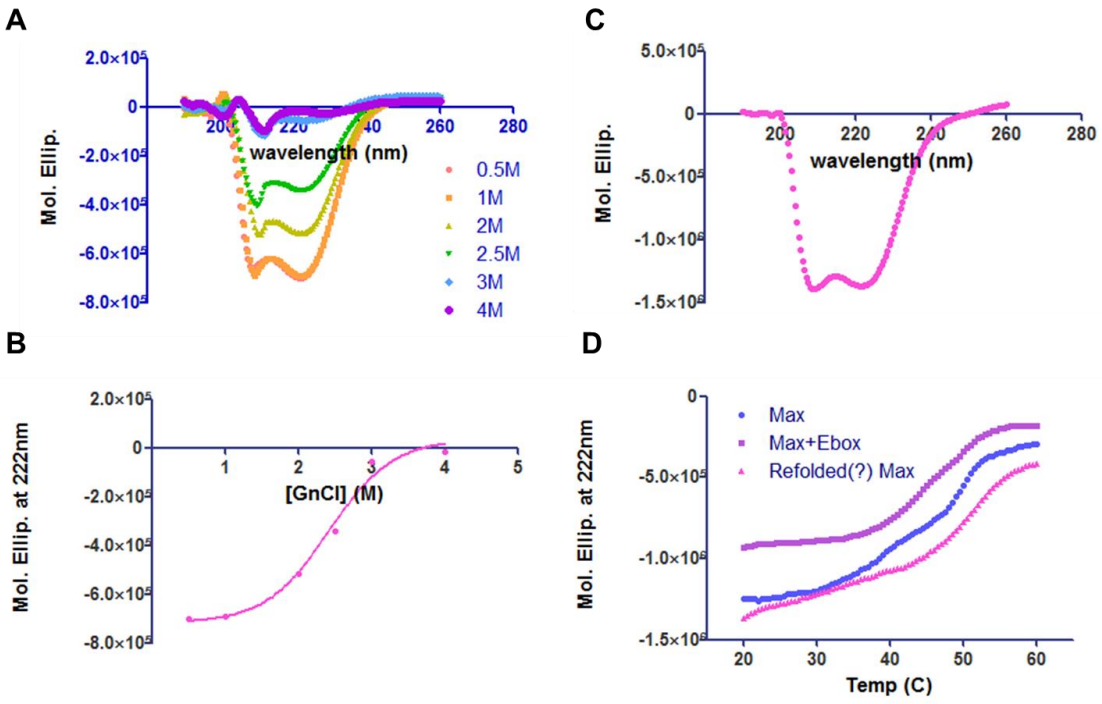


Figure 2.7. Denaturation and refolding of Max bHLH-Zip. (A)Purified Max was dissolved in increasing concentration of guanidinium hydrochloride; loss of secondary structure was monitored by CD. (B)Denaturation curve of Max. (C)CD spectrum of Max after refolding. (D)Thermo-denaturation curves of native and refolded Max.

result was one faced by many before – an incorrectly folded protein soluble in organic solvents or denaturants only. Occasionally denatured proteins could be refolded and re-solubilized under native conditions, typically done through dialysis, rapid dilution, or column-assisted refolding[10-16].

Since even "native" ACMax was prone to aggregation, we used it as a starting point to practice refolding. We had discussed of the use of Ni-NTA and size exclusion chromatography in an attempt to provide solid phase support for the conjugate during refolding. Rapid dilution was not met with much success, either. The next thing we tried was step-wise dialysis. First, denaturation curve of ACMax by increasing concentration of guanidinium hydrochloride was constructed using circular dichoism to monitor loss of secondary structure. The denatured protein was then dialyzed from 6M to 2.5M guanidinium hydrochloride, which was the point of inflection on the denaturation curve, and further dialyzed into PBS. The protein solution remained clear, and CD measurements showed more secondary structure, albeit different compared to the native protein.

The same conditions did not work, however, when applied to the conjugated product. As regular conditions and buffered failed, we conducted a screen of additives in buffer that have been reported to stabilize and promote folding of protein in a 94-well format, where denatured conjugate was rapidly diluted to different buffers. The readout was absorbance at 340nm measuring turbidity. Many conditions were screen in a combinatorial way, but no satisfyingly clear result was seen.

Topic Closing Remarks

We have created multiple versions of the SRD-bHLH-Zip fusion protein envisioned in Chapter I using both native chemical ligation and maleimide-cysteine conjugation; however, these semi-synthetic proteins have not been tested for hypothesized activities due to limitations of their physical properties. The knowledge gained from improving SRD affinity and solubility, as well as the methodologies established for ligation or conjugation, should be applicable to future efforts on the ultimate goal. Mainly, there are three directions in which improvements can be made: first, SRD could be further optimized either by scanning of more variations of hydrophobic residues, or by switching the current all-hydrocarbon staples with a newly developed amino-stapling system that generally incurs less hydrophobicity in product peptides (Hilinski and Verdine, unpublished data). Second, as mentioned in chapter I, other bHLH-Zip domains may be used in lieu of Max. This effort may be particularly worthwhile given Max's propensity to aggregate as observed. The new candidate need not be within the Myc/Max/Mad family – possession of comparable dimerization domain is good enough, since DNA recognition (basic region) could be engineered. Third, more soluble building blocks may enable conjugation or ligation under native conditions to avoid need of refolding.

Experimental Methods

Maleimide-peptide Synthesis

Maleimide-containing peptides were synthesized using standard Fmoc-SPPS conditions as described in Chapter I with a few additional steps. Before the beginning of the actual sequence, an Mmt-protected lysine was installed at the C-terminus followed by a linker (e.g.

PEG(2)). The peptide was then elongated and metathesis was performed after sequence completion. Afterwards, resin was subjected to treatment with 1% TFA, 4% TIS, and 95% DCM for 2 min and drained. The process was repeated with fresh reagents five to six times or until no longer orange color was observed, a signal of complete removal of the Mmt protecting group. SMCC (Pierce) was then coupled to the deprotected lysine side chain by addition of 3 equivalents of material and 5 equivalents of DIPEA, gently agitated under nitrogen for 4h. After SMCC coupling, the N-terminus of the peptide could be deprotected and labeled with FITC as usual. The purification process using HPLC was the same for maleimide-peptides as other stapled peptides.

Cysteine-Maleimide Conjugation

N-cysteine Max bHLH-Zip was expressed and purified as described in Chapter I. After Ni-NTA affinity purification, Max in Elution Buffer was precipitated with acetone at -20°C overnight. Pelleted Max was dissolved in Conjugation Buffer at a final concentration of 0.2mM and incubated at room temperature for 0.5h in order to first reverse the covalent modification by pyruvate. Then maleimide-containing peptide was added to the protein solution at a final concentration of 1mM. Conjugation was allowed to proceed overnight with gentle agitation at room temperature. Reaction completion was checked by LC/MS analysis, and the crude product was purified on reverse phase HPLC after desalting. Pooled fractions were dried using speedvac and lyophilized into powder form.

Conjugation Buffer
400mM o-methylhydroxylamine
6M guanidinium hydrochloride
150Mm sodium phosphate pH 7.8

Cloning, Expression, and Purification of ACMax

The ACMax construct was obtained by site-directed mutagenesis using Stratagene's QuickChange kit and alanine was inserted to the N-terminus before cysteine. ACMax was expressed and purified following the same protocol as described in Chapter I. After purification by Ni-NTA, it was noticed that the protein sample contained large amounts of DNA ($A_{260/280}$ = 1.8). Addition of DNase I into the Lysis Buffer did not eliminate the DNA effectively.

Pooled elution from Ni-NTA was then run on HighTrap heparin column (GE) and exchanged from Heparin Buffer A to Heparin Buffer B (high salt). This step was able to separate ACMax from DNA by disrupting their binding with high salt concentration. Pooled fractions from heparin column were loaded to Superdex75 gel filtration column in order to exchange back to a low salt buffer (Gel Filtration Buffer). Fractions from gel filtration were confirmed to be the correct, non-aggregated and DNA-free ACMax by SDS-PAGE and new $A_{260/280}$ readings.

Heparin Buffer A
20mM HEPES pH 7.5
300mM KCl
5% Glycerol
1mM DTT

Heparin Buffer B
20mM HEPES pH 7.5
2M KCl
5% Glycerol
1mM DTT

Gel Filtration Buffer
20mM HEPES pH 7.5
150mM KCl
5% Glycerol

Refolding of ACMax

Chemical Denaturation Curve: Purified ACMax was treated with increasing concentration of guanidinium hydrochloride. At each concentration, circular dichroism of the protein sample was measured and the molar ellipticity at 222nm was used as an indication of decrease in secondary structure. By plotting molar ellipticity at 222nm against the concentration of guanidinium hydrochloride and taking the point of inflection, the "critical" concentration of guanidinium hydrochloride at which half of the protein was denatured as found to be 2.5mM.

Refolding and Analysis of Thermo-stability of ACMax: Completely denatured ACMax (in 6M guanidinium hydrochloride) was refolded by stepwise dialysis: first into Gel Filtration Buffer with 2.5mM guanidinium hydrochloride, then 1M, then eventually achieving complete removal of guanidinium hydrochloride. Refolding was confirmed by circular dichoism. In addition, CD experiments were also used to the melting curve of ACMax and continuous recording of the spectrum over rising temperature range of 20 – 90°C.

References

1. Kleineweischede, R. & Hackenberger, C.P.R. Chemoselective peptide cyclization by traceless Staudinger ligation. *Angewandte Chemie-International Edition* **47**, 5984-5988 (2008).
2. Hackenberger, C.P. & Schwarzer, D. Chemoselective ligation and modification strategies for peptides and proteins. *Angew Chem Int Ed Engl* **47**, 10030-74 (2008).
3. Sletten, E.M. & Bertozzi, C.R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* **48**, 6974-98 (2009).
4. Kurtenbach, E. & Verjovski-Almeida, S. Labeling of a thiol residue in sarcoplasmic reticulum ATPase by pyrene maleimide. Solvent accessibility studied by fluorescence quenching. *J Biol Chem* **260**, 9636-41 (1985).
5. Ghosh, S.S., Kao, P.M., McCue, A.W. & Chappelle, H.L. Use of maleimide-thiol coupling chemistry for efficient syntheses of oligonucleotide-enzyme conjugate hybridization probes. *Bioconjug Chem* **1**, 71-6 (1990).
6. Cai, W., Zhang, X., Wu, Y. & Chen, X. A thiol-reactive 18F-labeling agent, N-[2-(4-18F-fluorobenzamido)ethyl]maleimide, and synthesis of RGD peptide-based tracer for PET imaging of alpha v beta 3 integrin expression. *J Nucl Med* **47**, 1172-80 (2006).
7. Slavica, A., Dib, I. & Nidetzky, B. Selective modification of surface-exposed thiol groups in Trigonopsis variabilis D-amino acid oxidase using poly(ethylene glycol) maleimide and its effect on activity and stability of the enzyme. *Biotechnol Bioeng* **96**, 9-17 (2007).
8. Pounder, R.J., Stanford, M.J., Brooks, P., Richards, S.P. & Dove, A.P. Metal free thiol-maleimide 'Click' reaction as a mild functionalisation strategy for degradable polymers. *Chem Commun (Camb)*, 5158-60 (2008).
9. Bernal, F., Tyler, A.F., Korsmeyer, S.J., Walensky, L.D. & Verdine, G.L. Reactivation of the p53 tumor suppressor pathway by a stapled p53 peptide. *Journal of the American Chemical Society* **129**, 2456-7 (2007).
10. Oganesyan, N., Kim, S.H. & Kim, R. On-column protein refolding for crystallization. *J Struct Funct Genomics* **6**, 177-82 (2005).
11. Langenhof, M., Leong, S.S., Pattenden, L.K. & Middelberg, A.P. Controlled oxidative protein refolding using an ion-exchange column. *J Chromatogr A* **1069**, 195-201 (2005).
12. Yazdanparast, R., Esmaeili, M.A. & Khodarahmi, R. Protein refolding assisted by molecular tube based alpha-cyclodextrin as an artificial chaperone. *Biochemistry (Mosc)* **71**, 1298-306 (2006).
13. Hagen, A.J., Hatton, T.A. & Wang, D.I. Protein refolding in reversed micelles. 1990. *Biotechnol Bioeng* **95**, 285-94 (2006).
14. Hackenberger, C.P., Chen, M.M. & Imperiali, B. Expression of N-terminal Cys-protein fragments using an intein refolding strategy. *Bioorg Med Chem* **14**, 5043-8 (2006).
15. Chow, M.K. et al. REFOLD: an analytical database of protein refolding methods. *Protein Expr Purif* **46**, 166-71 (2006).
16. Jungbauer, A. & Kaar, W. Current status of technical protein refolding. *J Biotechnol* **128**, 587-96 (2007).

Chapter III – Design, Synthesis, and Screening of a Library of Stapled Insulin Receptor Binding Peptides

Introduction

Diabetes is one of the most common and costly chronic diseases worldwide, and its prevalence has been growing at a faster pace in recent decades due to an exploding and aging global population, urbanization, and the obesity epidemic resulting from unhealthy diet and sedentary lifestyle. The United States, as with many other first world countries, has not fared better than poorer nations in either prevalence or societal impact of diabetes. In the latest CDC release of national diabetes statistics, 8.3% of the U.S. population is affected by either Type I or Type II diabetes. Healthcare expenditures directly imposed by the disease exceeded $218 billion annually as of 2011. Since diabetics are at elevated risk of other serious complications such as cardiovascular disease, stroke, hypertension, etc., quality of life is eroded further. Related medical costs and loss of productivity amount to additional economic burden and profound social problems. These facts underscore the urgency of deciphering diabetes at a molecular level as well as adapting advanced chemical toolkits so as to yield more effective therapeutics for controlling the disease.

Understanding the biology of diabetes requires capture of the motion picture that is the insulin signaling pathway. Binding of insulin to the transmembrane insulin receptor (IR)

induces conformational changes within the receptor, which result in auto-phosphorylation of intracellular β-subunits and activation of kinase activity of IR[1,2] and subsequent tyrosine phosphorylation of its substrate IRS proteins[3,4]. Much of the distal signaling involves activation of phosphatidylinositol 3'-kinase (PI3K) by docking of its p85 subunit to tyrosine-phosphorylated IRS-1 and IRS-2, which in turn leads to phosphorylation of protein kinase B (PKB/Akt) and downstream responses of glucose uptake and glycogen synthesis. Additionally, activated IRS-1 is also coupled the Ras pathway and up-regulation on gene expression[4-8]. Various kinds of disruption at the front end of this signaling pathway, e.g. lack of insulin secretion or insensitization of the receptor to insulin, mark the pathological origin of both Type I and Type II diabetes. Despite extensive effort over decades, the molecular mechanism of regular insulin binding and IR activation and of changes in disease states remains unclear.

As a member of the tyrosine kinase receptor family, the insulin receptor is a glycoprotein consisting of two α and two β subunits ($\alpha_2\beta_2$) covalently linked by disulfide bonds[9]. The IR, which is synthesized as a single chain proreceptor, undergoes glycosylation and dimerization in the endoplasmic reticulum followed by Furin cleavage into the α- and β-chains in the Golgi apparatus. The IR ectodomain is comprised of two full-length α subunits and the N-termini of two β subunits; the remaining C-termini of the β subunits contain the transmembrane (TM) domains and cytoplasmic tyrosine kinase (TK) domains[10,11]. Insulin binding determinants reside exclusively within the ectodomain, which consists of N-terminal two leucine rich repeat (L1 and L2) domains separated by a cystein-rich (CR) domain[12,13] as well as C-terminal three fibronectin type III (Fn0, Fn1, and Fn2) domains[14-16]. Additionally, an insertion domain residing in Fn1 contains a Furin cleavage site. Upon cleavage of the proreceptor, the α-α linkage is maintained by disulfide bonds between

Cys524 in each Fn0 domain[17] as well as those between Cys682, Cys683, and Cys685 in each insertion domain[18,19], while the α-β hetero-dimer is bridged by a single disulfide link between Cys647 in Fn1 and Cys872 in Fn2 domain[20] (Figure 3.1A).

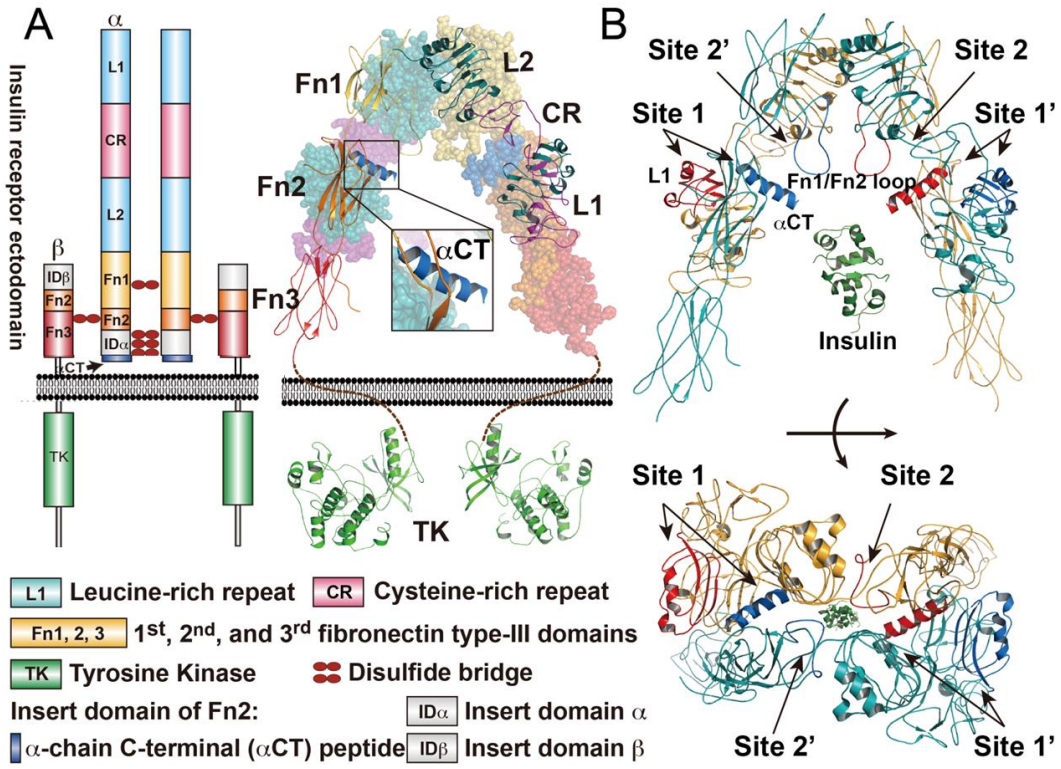


Figure 3.1. Representative domain structure of the insulin receptor. (A)Two α subunits and the N-termini of two β subunits form the ectodomain, whereas the C-terminal β subunits contain the transmembrane (TM) domains and tyrosine kinase (TK) domains. The ectodomain consists of the N-terminal leucine rich repeats and cysteine rich (L1-CR-L2) domains and three C-terminal fibronectin type III domains (Fn0, Fn1, and Fn2). α-CT resides at the C-terminus of the IR α-chain. (B) Insulin binding sites on IR. Site 1 is formed by the central β-sheets of L1 and αCT, and site 2 is the loop between Fn0 and Fn1.

Although insulin, the first peptide hormone visualized by x-ray crystallography[21], has been extensively explored structurally over the past 50 years, crystal structures of the L1-CR-L2 domain[22] and the whole ectodomain of IR[23,24] were not solved until 2006. Furthermore, neither of these constructs adopts a conformation conducive to high-affinity insulin binding.

Thus, high-resolution three-dimensional structural information of the insulin/IR binary complex continues to spark enormous interest. The current interaction model was obtained by indirect methods such as alanine scanning mutagenesis[25,26], photoaffinity labeling[27,28], chemical cross-linking[29,30], phage-displayed peptide probes[31], or monoclonal antibody probes[32-34]. Insulin binding to the dimeric holoreceptor is characterized generally by high binding affinity (pM range) and negative cooperativity[35,36]. It is proposed that there are two insulin binding sites on the IR: site 1, which contacts the classical binding surface of insulin[37,38], and site 2, which was predicted to bind the hexamer-forming face of insulin[39,40]. As shown in Figure 3.1B, site 1 on one monomer of the IR is close to site 2' on the second monomer, and binding of insulin to site 1 induces its subsequent binding to site 2'. This causes a conformational change of the IR ectodomain that leads to a reduction of distance between the two intercellular TK domains, thereby promoting autophosphorylation. Several lines of evidence have shown that binding site 1 on the IR is formed by the central β-sheet of the L1 domain[41,42] and the C-terminal α-subunit peptide (α-CT, aa704-aa719)[26,30]. Site 2 is likely to be located at the loop region between Fn0 and Fn1 since it faces site 1 of the other monomer in the dimeric structure of the IR ectodomain[24].

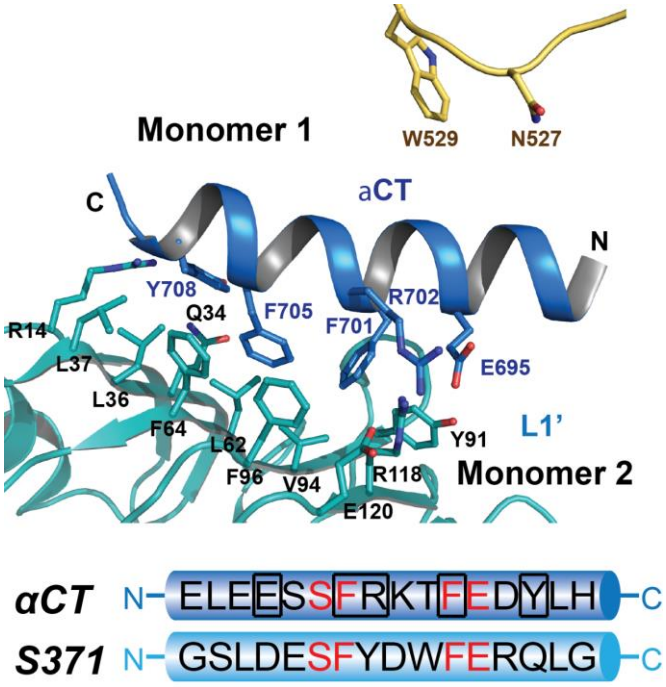


Figure 3.2. Structure of the a-CT segment of IR and sequence comparison with S371. Top, close-up view of interaction between αCT (blue) and L1 (teal). Bottom, sequence comparison of αCT and S371. Black square indicates residues engaged in L1 binding.

Insulin receptor binding peptides

had previously been evolved by phage display from randomized peptide libraries ranging in size from 20 to 40 amino acids.[31] Such primary binders were then subjected to defined mutagenesis procedures based on analysis of sequence motifs, consensus regions, and multiple independent hotspots on the receptor to yield secondary and tertiary libraries. Hits from all libraries were grouped to two major categories corresponding to two different binding sites on IR as determined using a competition ELISA assay. From these libraries came S371, a prototypical site 1 peptide with the highest affinity to the IR (k_d = 40 nM) and predicted to form a helix. Of particular interest is the similarity in sequence motif between S371 and the alpha subunit C-terminus (α-CT) of the IR (figure 3.2.), which had been speculated to be helical and play a role in ligand binding and receptor activation for quite some time. Several lines of evidence pointed to a model in which α-CT together with L1 of the other monomer forms an insulin biding site: 1) photoactivated cross-linking experiments[43-46] that associate α-CT with insulin at residues A3, A8 and B25, while insulin B24 can crosslink to L1, 2) mutagenesis[26,47] that showed Phe714 of α-CT is critical for ligand binding, and 3) study of mini-receptor[22,48,49] consisting of only L1-CR-L2 that does not bind insulin, to which a mere addendum of α-CT regains ligand binding with similar affinity to the soluble ectodomain. Excitingly, recent revelation of the α-C-terminal structure[24] confirmed its alpha-helical nature as well as binding to the beta-sheets L1 domain *in trans*. Although this piece of long-missing crystallography information closed the case of α-CT's place and associations in the ligand-free receptor, it raises more questions on the structural perturbations of ligand binding and interactions between α-CT, L1, and insulin. Furthermore, how does a surrogate peptide bind to the IR via a sequence- and perhaps structure-mimicry of α-CT, and could this surrogate ever be modified such that its binding affords a full transformation of receptor configuration resulting in activation much like the natural ligand? Thus it comes as no surprise that we select S371 as the lead compound for

enhancement with a hydrocarbon cross-link, for the goals of chemical optimization of the peptide and elucidation of its bound structure to the receptor. Through this practice we hope to arrive at compounds with greater potency as either agonist or antagonist to the IR which, beyond the obvious therapeutic potential, would also help unveil mechanism of IR activation at a greater detail than current knowledge provides.

Design and Synthesis of Stapled Insulin Receptor Binding Peptide Library (SIRB)

As discussed in previous chapters, the design of a stapled peptide library typically starts with a careful examination of the known ligand-bound structure of the target and selection of residues on the ligand (α-helix to be modified) that would endure replacement by non-natural amino acids with minimal cannibalization of natural interactions. A certain face of the helix away from the interface usually presents most sites suitable for modification. Such has been the logical thinking and historic holy grail in the burgeoning stapled peptide space, but recent studies prompt one to think outside the box. First, Phillips and colleagues at Pfizer reported stapled versions of a coactivator peptide binding to the estrogen receptor (ER) in 2011[50]. Importantly, crystal structure of the peptide-ER complex showed that the hydrocarbon crosslink is in fact part of the expanded interface, and actively engages hydrophobic interactions with the receptor that are unattainable to the unmodified peptide. Shortly after, structure of a previously published stapled P53 peptide bound to Mdm2 was solved, revealing similarly an active role of the staple in forming new contacts with the protein[51]. These discoveries add to the reasons for the dramatic increase in affinity oftentimes observed with these molecules – that beyond exerting physical constrain of the helix conformation, the staple itself could pick up additional dimensions that augment the existing protein-protein interactions. It is arguably a redeeming trait of the lengthy,

hydrophobic patch that is the crosslink; the grease ball that aggravates solubility issues to such problematic extents is not inherently without its merits (sometimes).

Thus, one may see the caveat of purposely avoiding placement of the crosslink at or near interacting residues or faces of the helix. The new "anything is possible" principle is, however, as exciting as it is hard to execute, for the sheer number of possibilities pose a significant challenge to the intellectually mundane but labor- and resource-consuming act of peptide synthesis. Furthermore, the screening element becomes more pronounced in a stapled peptide endeavor with larger library size, much like conventional fishing of small molecules, and the design part is diminished. Length of the original helix, as well as the resolve for how comprehensive the library should be, dictates the scale of the synthetic project. S371 spans 16 residues, a length that renders a comprehensive sampling of the hydrocarbon crosslink positions, as discussed below, elaborate yet still achievable. In light

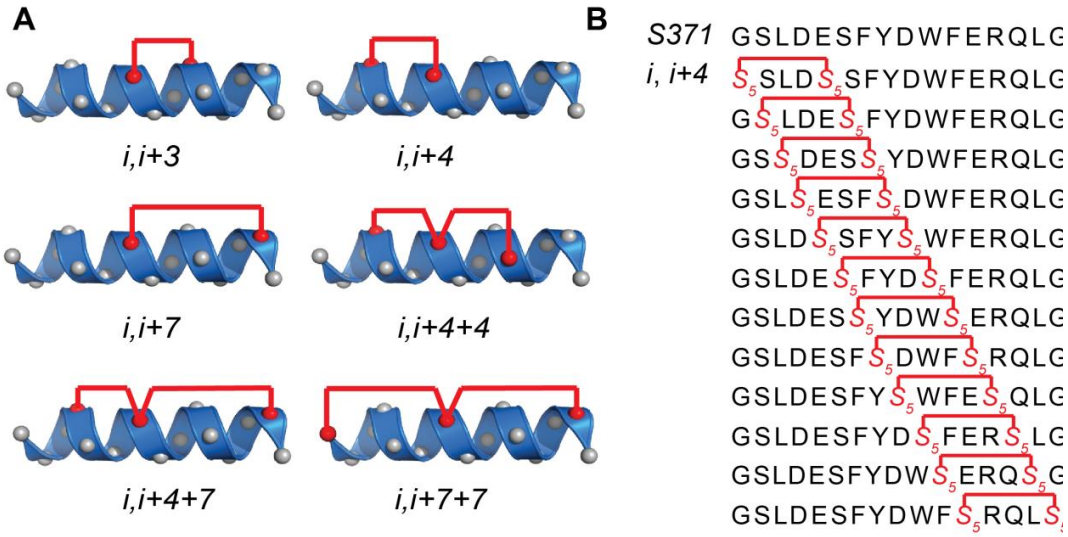


Figure 3.3. Schematic representation of all types of hydrocarbon crosslink implemented on peptide S371. (A) Six types of hydrocarbon crosslinks featuring single staple (*i, i+3; i, i+4; i, i+7*) and stitching (*i, i+4+4; i, i+4+7, i, i+7+7*.). Red spheres represent non-natural amino acids. (B)Example of staple "scanning" through the S371 sequence for *i, i+4*.

of previous examples, we envision that adding a hydrocarbon crosslink on S371 will impart important properties: first, greater secondary structure that will lead to higher target affinity; and second, resistance to proteolytic degradation, higher bioavailability, and longer serum half-life will improve pharmacokinetics on the therapeutic perspective. A final, more ambitious trait that may be actualized by this chemical modification is that the hydrophobic nature and lengthy stretch of the cross-link may enable participation in Van der Waals interactions with the FnIII-1/FnIII-2 loop at site 2, engaging an interface not accessible to an unmodified site 1 peptide. As such, it is of particular interest to include as many staple sites as possible in this case, given that the helix alone is known to solicit only a limited set of interactions. In order to fully explore the physicochemical space reachable by the hydrocarbon cross-link, we have adopted a comprehensive scanning approach in creating the peptide library. Distinct combinations of α, α-disubstituted amino acids are used to form stapled insulin receptor binding (SIRB) peptides with cross-links of various length and composition (Figure 3.3). All seven types of hydrocarbon crosslinks that our laboratory has developed to date are included in the library design: *i, i+3; i, i+4; i, i+7; i, i+4+4; i, i+4+7; i, i+7+4;* and *i, i+7+7*. In each series, "i" starts at the C-terminus of the peptide and moves forth until the other end of the crosslink reaches the N-terminus. This way all possible positions for staple incorporation can be sampled. Stapled peptides were made by solid phase synthesis using standard Fmoc chemistry as described previously[52].

Screening of SIRB library and Validation of Hit Peptides

The importance of a robust and reliable screening system is paramount in such an all-out expedition. A direct read-out of affinity to the receptor is highly desired, and the assay to be relatively scalable to accommodate throughput of the library. Here again implementation of routine practice – that is, recombinantly obtaining the target protein and quantitatively determining the K_d of each peptide via fluorescence polarization, isothermal titration calorimetry, or other binding assays – encounters unusual impediment, because the insulin receptor protein is, in short, hard to get a hold of, especially for large enough quantities to conduct the screening (will be addressed in detail in Chapter 4). Drawing example from the literature[31,53-55], the next-best measurement is of activity of the peptide upon engaging the receptor (agonistic/antagonistic effects, etc.). To this end, changes in the insulin signaling cascade in the presence of peptide can be monitored by cell-based assays that circumvent the need of purified IR protein.

As mentioned before, the primary signal relay that initiates at the insulin receptor is phosphorylation of the protein itself and downstream effectors or "nodes". One of the most commonly assessed nodes in the IR pathway is PKB/Akt1, a growth-promoting, anti-apoptotic protein responsible for transmitting the mitogenic and metabolic effects of insulin[56-60]. We therefore wished to evaluate the activity of SIRB peptides by measuring the changes in phosphorylation level of Akt1. The first cell line we selected to use for screening is HepG2, human liver cancer cells which express high levels of IR and Akt1[61-63]. For initial screening, each peptide in the library was tested for both antagonistic and agonistic effects using a sandwich ELISA assay kit (Cell Signaling) by measuring phosphorylation of Akt1 at Ser473 performed on HepG2 cells. Two measurements were made on each compound: 1) treatment of cells with insulin and peptide together to reveal antagonism on the receptor; 2)

treatment with peptide alone to reveal agonistic effects. We first validated the assay with unmodified S371 and established the starting point activity: at 10 μM it demonstrated 20% reduction of the maximal phosphorylation level caused by insulin, and no trace of any increase in phosphorylation on its own. This observation is in slight contradiction with published agonistic attribute based on increase in kinase activity of IR in vitro[31]. Nevertheless, we do note that although S371 has a high affinity to IR, it can barely induce downstream effects such as lipogenesis in rat adipocytes, which may reconcile with the result of no increase in downstream Akt phosphorylation. With respect to our stapled peptides, one or two peptides emerged as antagonists in each crosslink series. The hits and their corresponding reduction of Akt phosphorylation at 10 μM are summarized in Figure 3.4. Intriguingly, one of these peptides, SIRB-D2, exhibited significant agonistic effects at a concentration of 100 μM, leading up to 60% of the maximal insulin-induced level (Figure 3.4B). We hypothesize that at such a high concentration, the SIRB-D2 peptide may yield a

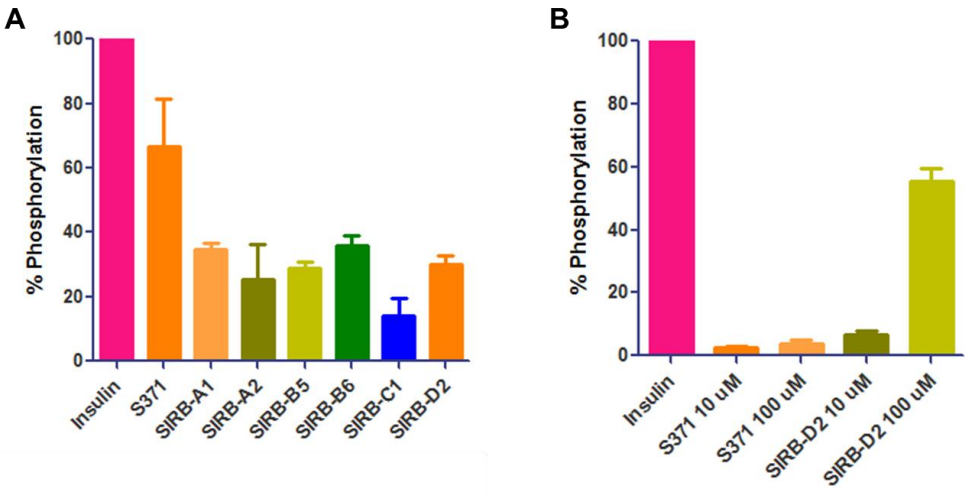


Figure 3.4. SIRB antagonism and agonism. Effects of SIRB on insulin signaling were measured by ELISA specific for Akt-S473 phosphorylation. (A)HepG2 cells were treated with 10 μM peptide in the presence of 50 nM insulin. In each series, there emerged SIRB peptides that exhibit greater than a 60% reduction of the insulin-induced signal. (B)HepG2 cells were treated with 10 or 100 μM peptide in the absence of insulin. SIRB-D2 at 100 μM exhibits agonism equal to approximately 60% of the insulin-induced signal. All readouts (in RLU) were normalized to the signal by 50 nM insulin alone with vehicle as baseline.

non-covalent dimer equivalent capable of engaging both binding sites on IR as insulin does, which still needs to be further demonstrated by complex structure of peptide bound IR ectodomain. No similar agonistic effect was observed with any other compound in the library.

The choice of HepG2 cells and Akt1 as a first line of screening was based on the ease of access: the cell line was commercially available and the antibodies of Akt are widely used and very well validated. But we needed a more direct assessment of the IR phosphorylation in order to be confident on the peptides' effects on distal signaling stated above. Furthermore, Akt1 is a node shared by many crisscrossing cellular circuits such as AMPK, mTOR, ErbB/HER pathways, etc.[56,64-66] Therefore off-target activity could not be excluded from the possible explanations for the observed effects. We were fortunate to have received in the nick of time a generous gift from Professor Morris White's lab at Boston Children's hospital - a stably transfected CHO-IR cell line from which phosphorylation of IR could be singled out with ease at the expression level of 10,000 receptors per cell. Antibodies that recognize the phosphorylated receptor cannot distinguish IR from insulin-like growth

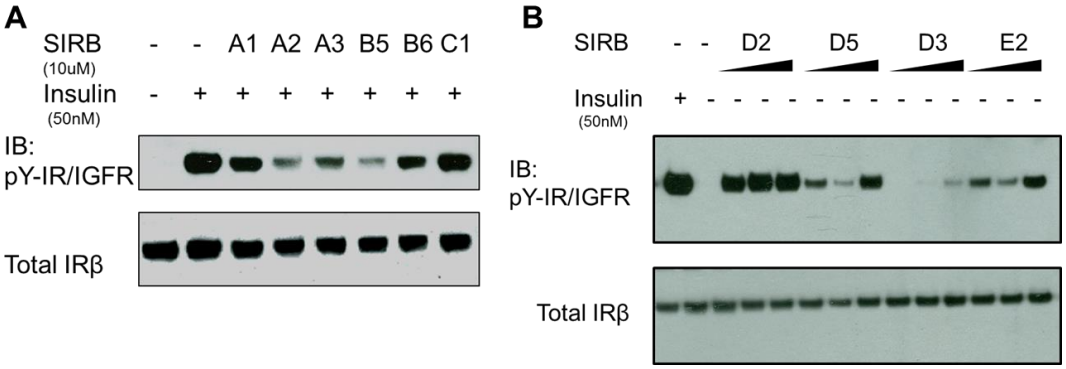


Figure 3.5. Validation of SIRB peptides. Effects of SIRB peptides on Phospho-IR level were detected by western blot. (A)CHO-IR cells were treated with 10 μM peptides in the presence of 50 nM insulin. (B) CHO-IR cells were treated with increasing concentration of peptide in the absence of insulin. Results are shown in comparison of stimulation by 50nM insulin. Total IR levels were measured as control.

factor 1 receptor (IGF1R), so the fact that little to none IGF1R is expressed in this cell line was quite convenient. The entire SIRB library was again screened for auto-phosphorylation of IR, with blotting of total IR-β as an internal control of loading. The results are aligned with previous observations on Akt1 to a satisfactory degree. On the antagonistic perspective, all of the pre-identified peptides that reduced phosphorylation signals of Akt1 in the presence of insulin also demonstrated similar effects on the IR, and no antagonists of IR alone were found that had not surfaced from the Akt1 ELISA (figure 3.5A) As for agonistic activity in the absence of insulin, there were surprises (figure 3.5B). Peptide SIRB-D2 indeed induced phosphorylation of IR in a dose-response manner, as expected from its effect on Akt1. There were, however, more peptides that are able to induce IR auto-phosphorylation but had seemingly no observable effect on Akt1: SIRB-D5 and SIRB-E2. This curious discrepancy may be indicative of the peptides' different potencies, i.e. how far down the pathway the signal is transmitted depends on the intensity and duration of the original event at the receptor.

We also quantified the dose-response behavior of the most active antagonist SIRB-B5 on inhibiting insulin-induced autophosphorylation of IR and phosphorylation of Akt with ELISA (figure 3.6). It was observed that SIRB-B5 exerts similar concentration-dependent antagonistic effects on both IR and Akt, with an IC50 value of 1.8 μM. Interestingly, at the maximal dose of 100 μM peptide, autophosphorylation of IR is reduced to about 6% of control, whereas Akt phosphorylation is reduced to a lesser extent, 16%. Again, a possible explanation for the differential plateau values is that phosphorylation of Akt1 is controlled by many other pathways as well and thus difficult to abolish completely.

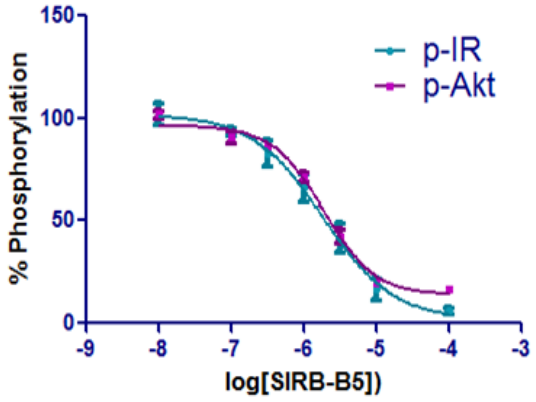


Figure 3.6. Dose-response of antagonist SIRB-B5. CHO-IR cells were treated with 50 nM insulin and increasing concentrations of SIRB5. ELISA assays of phosphor-IR Y1150/1151 and phosphor-Akt S473 were performed with cell lysates.

After using these independent measurements of phosphorylation levels as two rounds of screening and obtaining consistent results (for the most part), we felt confident of the peptides identified as either agonists or antagonists of the IR. In addition, circular dichroism revealed an increase in helicity of active SIRB peptides compared to unmodified S371 (Figure 3.7). Interestingly, peptide A2, one of the antagonists, shows an intense minimum of molar ellipticity around 208nm, characteristic of a 3_{10} helix. None of the peptides in other crosslink series exhibits such structural trait. As we move down the lists of series, i. e. going from relatively short and simple to longer and more constrained crosslinks, we observe the general trend of increasing helical character (compare B5, C1 and D2). Such results are in accordance with our general experience with stapled peptides, whose secondary structure is usually positively correlated with staple length and rigidity.

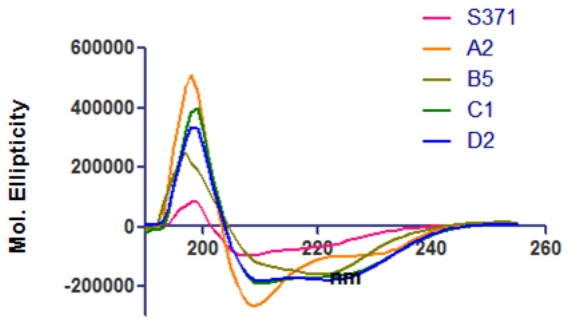


Figure 3.7. Secondary structure of S371 and select SIRB hits. CD spectroscopy of active SIRB peptides in each cross-link series illustrates the increase in helicity brought by synthetic modification.

Figure 3.8 shows a synthesis of the findings this far. Sequences of peptide hits are shown grouped in the two

categories of antagonists and agonists. Two points should be made here: 1) Although one or two peptides emerge from each type of hydrocarbon staple series as hits, all of them regardless of series or category contain the non-natural amino acids for stapling at converging sites, which if traced back to the original S371 sequence localize to only eight out a total of 16 residues. Structurally speaking, these sites concentrate on a particular face of the helical wheel, and we see no peptides that feature the staple on the opposite face being particularly active. 2) Though all hits share said localization of staple positions, only the longest crosslinks seem to confer any agonistic effects (SIRB D2, D5, and E5 are i+4+7 and i+7+7). Coinciding with this result is the trend of increasing helicity discussed above; thus it appears that a more structurally constrained peptide is more prone to antagonize the receptor. Yet as mentioned in the introduction, longer staples are themselves source of contact surface in addition to simply rigidifying the peptide. The question of whether the staples in the agonist peptides are able to pick up IR residues either nearby on the L1 domain or from the loop in the junction of Fn0/FnI domains of the purported Site 2 is both highly relevant and difficult to answer without direct structural information. A third possibility is that peptides of the later series, whether by virtue of being more structured or more hydrophobic thanks to their crosslinks, experience a higher proclivity of self-dimerization or oligomerization when put in contact with the receptor, and therefore engaging the two sites on IR and causing conformational changes much the way insulin does. Figure 3.10B attempts to shed light to the currently structure-less situation by modeling S371 based on the known position of IR αCT, using conserved residues as anchor points. The eight residues where staple "hotspots" were found are highlighted. If the hypothesis that S371 and thus the SIRB peptides mimic αCT in binding to the L1 domain is correct, and that the orientation of the peptide is as predicted, then it is no surprise that the "hotspots" are where they are: first, a staple placed there avoid clashing with the L1 surface directly;

second, extending on the back of the helix away from L1, the staple is positioned to either reach out for contact with Site 2 or serve as a scaffold for self-assembly via hydrophobic interactions, i.e. the dimer- or oligomerization scenario stated above.

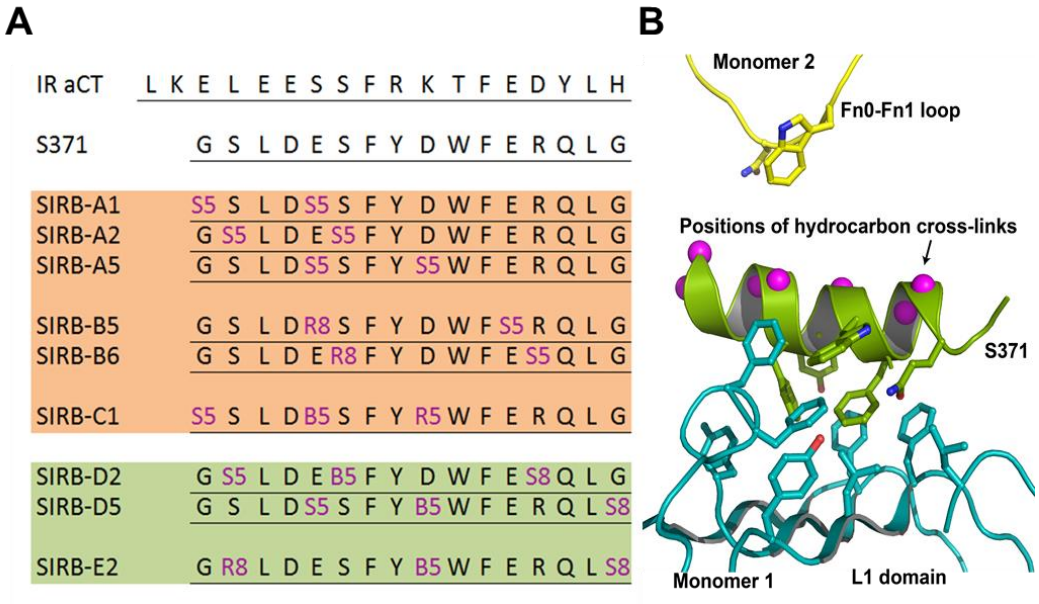


Figure 3.8. Summary of active SIRB peptides. (A) Sequences of antagonists (shaded in salmon) and agonists (shaded in green). (B) The interaction model between S371 (green) and L1 domain (insulin binding site 1 on IR, teal) of the IR. Staple hotspots are shown in magenta spheres. Residues involved in direct interactions between S371 and the IR as well as residues in Fn0/Fn1 loop (insulin binding site 2 on IR, in wheat) that may interact with staples are shown in stick rendering.

In summary, screening of SIRB library by ELISA of Akt1 phosphorylation identified peptides that either antagonize insulin signaling or agonize the IR pathway in the absence of insulin. Independent screening of receptor auto-phosphorylation by Western blot confirms the activity of hit peptides comes at least in part from directly interacting with IR. Secondary structure measurements and modeling of the peptide sequence based its alleged object of imitation – IR αCT- help rationalize the relations between crosslink chemistry and observed activity of the peptides, but more questions are raised along the way. How can the structural model be supported before direct evidence of the complex structure becomes available? If the hypothesis holds, what are the broader ramifications on the peptides' cellular activities,

especially their selectivity or promiscuity with respect to other receptor tyrosine kinases? Furthermore, are there ways to optimize the potency of the active peptides? These are the questions that Chapter 4 seeks to address.

Experimental Methods

Peptide Synthesis

SIRB Library peptides were synthesized using the standard Fmoc-SPPS protocol and purified on reverse-phase HPLC as described in Chapter I.

Cell Culture

HepG2 cells (ATCC) were maintained in complete growth medium consisting of DMEM, 10% fetal bovine serum (FBS) and streptomycin/penicillin. Cells were passaged when reaching greater than 80% confluence (on average 3-4 days) and split 1:6 per passage.

CHO-IR cells (from Prof. Morris White at Boston Children's Hospital) were maintained in complete growth medium consisting of Ham's F-12, 10% FBS and streptomycin/penicillin. Cells were passaged when reaching greater than 80% confluence (on average 2-3 days) and split 1:10 per passage.

Phospho-Akt1 ELISA

Preparation: insulin stock solution at 50μM (0.3mg/ml or 30mg/100ml). To make stock: dissolved 10mg in 1 ml 0.01N HCl, heat at 37C to dissolve, then added to 32mL PBS, sterile filtered and stored at -20C. Working stock would stay in good quality for 1 month at 4°C.

HepG2 cells grown to 80% confluent in 10cm plates or 6-well plates were serum-starved in DMEM (penicillin/streptomycin added) overnight. On the next day, the following solutions

were prepared: 50nM insulin in complete growth medium (from 50μM stock), 10μM or 100μM peptide in complete growth medium (from 10mM DMSO stock), and vehicle (the highest concentration of DMSO used in peptide samples). Old media was removed and cells were treated with appropriate compounds. For measurement of antagonism, the vehicle was used to create baseline, and the 50nM insulin solution was used to give maximal amount of signal (100% phosphorylation). Each of the peptides was given in combination with 50nM insulin. For measurement of agonism, the baseline and maximal level of phosphorylation were obtained the same way. Each of the peptides was given along in the absence of insulin. Cells were incubated at 37°C under 5% CO_2 for 15 min for antagonism and 30 min for agonism.

At the end of treatment, media was aspirated and cells were washed with ice cold PBS, and subsequently lysed using Lysis Buffer (from 10X Lysis Buffer, Cell Signaling; 1mM PMSF added prior to use). Cells were incubated in Lysis Buffer plates/wells in the cold room with gentle rocking for 10 min, then subjected to quick sonication. The cell lysate was then centrifuged at 12,000rpm for 10 min in the cold room. The clarified lysate was collected and aliquots were flash frozen in liquid nitrogen and stored at -80°C until use.

For the phospho-Akt1 ELISA assay, clarified lysate was first quantified using the standard BCA method, and all samples were adjusted to have the same total protein concentration. All samples were then diluted with Sample Diluent (from phospho-Akt1 S473 ELISA kit Cell Signaling) at 1:1, and added to antibody-coated microwells. Microwells were sealed firmly with tape and incubated at 4°C overnight.

The following day, wells were drained and washed four times with 1X Wash Buffer. 50ul detection antibody was added to each well, sealed with tape, and incubated 1h at room temperature. Contents were discarded and wells were washed with 1X Wash Buffer four

times. 50ul HRP-conjugated secondary antibody was to each well, seale with tape, and incubated for 30 minutes at room temperature. Contents were discarded and wells were washed again four times. Luminol/Enhancer solutions were mixed 1:1, and 50ul was added to each well. Chemiluminescence was read under RLU mode in the Hewlett-Packard SpectraMax Plate Reader at 425nM within 1-10 minutes of addition of substrate.

Immunoblotting of IR auto-phosphorylation

CHO-IR cells were seeded in 24-well plates and grown in Ham's F-12, 10% FBS medium to >80% confluent (in ~48 hours). The cells are starved in Ham's F-12 medium without FBS for two hours prior to experiment.

For agonistic studies, cells were incubated in fresh Ham's F-12 without FBS (negative control and baseline for normalization) or the same medium with 50nM insulin (positive control and 100% activity for normalization). Experimental wells were incubated in the same medium containing various concentrations of peptide (typically 1uM, 10uM, and 100uM). All cells were incubated at 37C for 15 min or 30 min.

For antagonistic studies, the experiment could be done with or without pre-treatment with peptides. In the case without, cells were incubated in fresh Ham's F-12 without FBS (negative control and baseline for normalization) or the same medium with 50nM insulin (positive control and 100% activity for normalization). Experimental wells were incubated in the same medium containing various concentrations of peptide (typically 1uM, 10uM, and 100uM) AND 50 nM insulin. All cells were incubated at 37C for 15 min. In the case with pre-treatment, experimental wells are pre-incubated with various concentration of peptide (typically 1uM, 10uM, and 100uM) at 37C for 30 min before insulin (final concentration

50nM) was added. In the case with pre-treatment, all is the same except the experimental groups were pre-incubated with peptide for 30 min before the addition of insulin.

At the end of incubation all medium is removed and cells are washed once with ice-cold PBS. The cells will remain on ice from this point on. Ice-cold lysis buffer is added (50-100uL/well for 24-well plates).

Composition of lysis buffer:

-SDS loading buffer, from 4X stock (To make 10 mL of 4x stock: 2.0 ml 1M Tris-HCl pH 6.8, 0.8 g SDS, 4.0 ml 100% glycerol, 0.4 ml 14.7 M β-mercaptoethanol, 1.0 ml 0.5 M EDTA, 8 mg bromophenol Blue; sonicate to dissolve)

-PMSF, 1uM, from 100X stock in ethanol

-Protease inhibitor 1 and phosphatase inhibitors 2 & 3 from Sigma Aldrich, all from 100X

-5% beta-mercaptoethanol

-Fill up volume to 1X with TBS, from 10X stock, Mediatech

(PMSF, protease inhibitors, and BME were added right before experiment)

Cells are lysed in wells for greater than 5 minutes then collected by cell scrapers or scraping gently with pipet tips. The lysates were boiled for 15 minutes. They could be directly loaded onto SDS-PAGE gel or stored at -80°C.

Western Blot:

SDS-PAGE gel is transferred to nitrocellulose or PVDF membrane (PVDF must be primed with methanol before use) for 100 min at 33V in the cold room.

The membrane is washed briefly in TBS-T (1X TBS, 0.1% Tween-20) and incubated in blocking buffer (5% BSA in TBS-T, filtered) for 1h at room temperature.

The membrane is washed 2X and incubated with primary antibody (Phospho-IGF-I Receptor β (Tyr1135/1136)/Insulin Receptor β (Tyr1150/1151) (19H7) Rabbit mAb Cell Signaling #3024), 1:300 in 3% BSA in TBS-T) overnight at 4C.

The membrane is washed four times, 5-10 minutes each in TBS-T, then incubated with secondary antibody (Anti-rabbit IgG, HRP-linked Antibody Cell Signaling #7074) 1:3000 in 3% BSA in TBS-T) for 1h at room temperature. The membrane was washed four times for 5-10 min. The chemiluminescence is developed by applying a mixture of Pico-level substrate and 1% Femto-level substrate (Thermo) to the blot.

Subsequently, the blot can be stripped and reblot for total IR (Insulin Receptor β (4B8) Rabbit mAb Cell Signaling #3025) in the same manner.

References

1. Hubbard, S.R., Wei, L., Ellis, L. & Hendrickson, W.A. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372**, 746-54 (1994).
2. Hubbard, S.R. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J* **16**, 5572-81 (1997).
3. Sun, X.J. et al. Structure of the Insulin-Receptor Substrate Irs-1 Defines a Unique Signal Transduction Protein. *Nature* **352**, 73-77 (1991).
4. Saltiel, A.R. & Kahn, C.R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799-806 (2001).
5. White, M.F. & Yenush, L. The IRS-signaling system: A network of docking proteins that mediate insulin and cytokine action. *Protein Modules in Signal Transduction* **228**, 179-208 (1998).
6. Taniguchi, C.M., Emanuelli, B. & Kahn, C.R. Critical nodes in signalling pathways: insights into insulin action. *Nature Reviews Molecular Cell Biology* **7**, 85-96 (2006).
7. Taniguchi, C.M. et al. Phosphoinositide 3-kinase regulatory subunit p85alpha suppresses insulin action via positive regulation of PTEN. *Proc Natl Acad Sci U S A* **103**, 12093-7 (2006).
8. Taniguchi, C.M., Ueki, K. & Kahn, R. Complementary roles of IRS-1 and IRS-2 in the hepatic regulation of metabolism. *J Clin Invest* **115**, 718-27 (2005).
9. Siddle, K. et al. Specificity in ligand binding and intracellular signalling by insulin and insulin-like growth factor receptors. *Biochem Soc Trans* **29**, 513-25 (2001).
10. Marino-Buslje, C., Martin-Martinez, M., Mizuguchi, K., Siddle, K. & Blundell, T.L. The insulin receptor: from protein sequence to structure. *Biochem Soc Trans* **27**, 715-26 (1999).
11. De Meyts, P. The insulin receptor: a prototype for dimeric, allosteric membrane receptors? *Trends Biochem Sci* **33**, 376-84 (2008).
12. Bajaj, M., Waterfield, M.D., Schlessinger, J., Taylor, W.R. & Blundell, T. On the tertiary structure of the extracellular domains of the epidermal growth factor and insulin receptors. *Biochim Biophys Acta* **916**, 220-6 (1987).
13. Ward, C.W., Hoyne, P.A. & Flegg, R.H. Insulin and epidermal growth factor receptors contain the cysteine repeat motif found in the tumor necrosis factor receptor. *Proteins* **22**, 141-53 (1995).
14. Ward, C.W. Members of the insulin receptor family contain three fibronectin type III domains. *Growth Factors* **16**, 315-22 (1999).
15. Marino-Buslje, C., Mizuguchi, K., Siddle, K. & Blundell, T.L. A third fibronectin type III domain in the extracellular region of the insulin receptor family. *FEBS Lett* **441**, 331-6 (1998).
16. Mulhern, T.D., Booker, G.W. & Cosgrove, L. A third fibronectin-type-III domain in the insulin-family receptors. *Trends Biochem Sci* **23**, 465-6 (1998).
17. Schaffer, L. & Ljungqvist, L. Identification of a disulfide bridge connecting the alpha-subunits of the extracellular domain of the insulin receptor. *Biochem Biophys Res Commun* **189**, 650-3 (1992).
18. Sparrow, L.G. et al. The disulfide bonds in the C-terminal domains of the human insulin receptor ectodomain. *J Biol Chem* **272**, 29460-7 (1997).
19. Lu, K. & Guidotti, G. Identification of the cysteine residues involved in the class I disulfide bonds of the human insulin receptor: properties of insulin receptor monomers. *Mol Biol Cell* **7**, 679-91 (1996).

20. Cheatham, B. & Kahn, C.R. Cysteine 647 in the insulin receptor is required for normal covalent interaction between alpha- and beta-subunits and signal transduction. *J Biol Chem* **267**, 7108-15 (1992).
21. Blundell, T.L. et al. Atomic positions in rhombohedral 2-zinc insulin crystals. *Nature* **231**, 506-11 (1971).
22. Lou, M. et al. The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity. *Proc Natl Acad Sci U S A* **103**, 12429-34 (2006).
23. McKern, N.M. et al. Structure of the insulin receptor ectodomain reveals a folded-over conformation. *Nature* **443**, 218-21 (2006).
24. Smith, B.J. et al. Structural resolution of a tandem hormone-binding element in the insulin receptor and its implications for design of peptide agonists. *Proc Natl Acad Sci U S A* **107**, 6771-6 (2010).
25. Williams, P.F., Mynarcik, D.C., Yu, G.Q. & Whittaker, J. Mapping of an NH2-terminal ligand binding site of the insulin receptor by alanine scanning mutagenesis. *J Biol Chem* **270**, 3012-6 (1995).
26. Mynarcik, D.C., Yu, G.Q. & Whittaker, J. Alanine-scanning mutagenesis of a C-terminal ligand binding domain of the insulin receptor alpha subunit. *Journal of Biological Chemistry* **271**, 2439-2442 (1996).
27. Yip, C.C. et al. Localization of the insulin-binding site to the cysteine-rich region of the insulin receptor alpha-subunit. *Biochem Biophys Res Commun* **157**, 321-9 (1988).
28. Fabry, M. et al. Detection of a new hormone contact site within the insulin receptor ectodomain by the use of a novel photoreactive insulin. *J Biol Chem* **267**, 8950-6 (1992).
29. Waugh, S.M., DiBella, E.E. & Pilch, P.F. Isolation of a proteolytically derived domain of the insulin receptor containing the major site of cross-linking/binding. *Biochemistry* **28**, 3448-55 (1989).
30. Kurose, T. et al. Cross-linking of a B25 azidophenylalanine insulin derivative to the carboxyl-terminal region of the alpha-subunit of the insulin receptor. Identification of a new insulin-binding domain in the insulin receptor. *J Biol Chem* **269**, 29190-7 (1994).
31. Pillutla, R.C. et al. Peptides identify the critical hotspots involved in the biological activation of the insulin receptor. *Journal of Biological Chemistry* **277**, 22590-22594 (2002).
32. Taylor, R., Soos, M.A., Wells, A., Argyraki, M. & Siddle, K. Insulin-like and insulin-inhibitory effects of monoclonal antibodies for different epitopes on the human insulin receptor. *Biochem J* **242**, 123-9 (1987).
33. Zhang, B. & Roth, R.A. A region of the insulin receptor important for ligand binding (residues 450-601) is recognized by patients' autoimmune antibodies and inhibitory monoclonal antibodies. *Proc Natl Acad Sci U S A* **88**, 9858-62 (1991).
34. Prigent, S.A., Stanley, K.K. & Siddle, K. Identification of epitopes on the human insulin receptor reacting with rabbit polyclonal antisera and mouse monoclonal antibodies. *J Biol Chem* **265**, 9970-7 (1990).
35. de Meyts, P., Roth, J., Neville, D.M., Jr., Gavin, J.R., 3rd & Lesniak, M.A. Insulin interactions with its receptors: experimental evidence for negative cooperativity. *Biochem Biophys Res Commun* **55**, 154-61 (1973).
36. De Meyts, P. The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling. *Diabetologia* **37 Suppl 2**, S135-48 (1994).

37. Hua, Q.X., Shoelson, S.E., Kochoyan, M. & Weiss, M.A. Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* **354**, 238-41 (1991).
38. Ludvigsen, S., Olsen, H.B. & Kaarsholm, N.C. A structural switch in a mutant insulin exposes key residues for receptor binding. *J Mol Biol* **279**, 1-7 (1998).
39. Glendorf, T., Sorensen, A.R., Nishimura, E., Pettersson, I. & Kjeldsen, T. Importance of the solvent-exposed residues of the insulin B chain alpha-helix for receptor binding. *Biochemistry* **47**, 4743-51 (2008).
40. Schaffer, L. A model for insulin binding to the insulin receptor. *Eur J Biochem* **221**, 1127-32 (1994).
41. Huang, K. et al. How insulin binds: the B-chain alpha-helix contacts the L1 beta-helix of the insulin receptor. *J Mol Biol* **341**, 529-50 (2004).
42. Mynarcik, D.C., Williams, P.F., Schaffer, L., Yu, G.Q. & Whittaker, J. Identification of common ligand binding determinants of the insulin and insulin-like growth factor 1 receptors. Insights into mechanisms of ligand binding. *J Biol Chem* **272**, 18650-5 (1997).
43. Wan, Z.L. et al. Diabetes-associated mutations in human insulin: Crystal structure and photo-cross-linking studies of A-chain variant insulin Wakayama. *Biochemistry* **44**, 5000-5016 (2005).
44. Huang, K. et al. The A-chain of insulin contacts the insert domain of the insulin receptor. Photo-cross-linking and mutagenesis of a diabetes-related crevice. *J Biol Chem* **282**, 35337-49 (2007).
45. Wan, Z.L. et al. Enhancing the activity of insulin at the receptor interface: Crystal structure and photo-cross-linking of A8 analogues. *Biochemistry* **43**, 16119-16133 (2004).
46. Kurose, T. et al. Cross-Linking of a B25 Azidophenylalanine Insulin Derivative to the Carboxyl-Terminal Region of the Alpha-Subunit of the Insulin-Receptor - Identification of a New Insulin-Binding Domain in the Insulin-Receptor. *Journal of Biological Chemistry* **269**, 29190-29197 (1994).
47. Mynarcik, D.C., Williams, P.F., Schaffer, L., Yu, G.Q. & Whittaker, J. Analog binding properties of insulin receptor mutants - Identification of amino acids interacting with the cooh terminus of the B-CHAIN of the insulin molecule. *Journal of Biological Chemistry* **272**, 2077-2081 (1997).
48. Garrett, T.P.J. et al. Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. *Nature* **394**, 395-399 (1998).
49. Kristensen, C., Wiberg, F.C., Schaffer, L. & Andersen, A.S. Expression and characterization of a 70-kDa fragment of the insulin receptor that binds insulin - Minimizing ligand binding domain of the insulin receptor. *Journal of Biological Chemistry* **273**, 17780-17786 (1998).
50. Phillips, C. et al. Design and Structure of Stapled Peptides Binding to Estrogen Receptors. *Journal of the American Chemical Society* **133**, 9696-9699 (2011).
51. Baek, S. et al. Structure of the Stapled p53 Peptide Bound to Mdm2. *Journal of the American Chemical Society* **134**, 103-106 (2012).
52. Kim, Y.W., Grossmann, T.N. & Verdine, G.L. Synthesis of all-hydrocarbon stapled alpha-helical peptides by ring-closing olefin metathesis. *Nature Protocols* **6**, 761-771 (2011).
53. Schaffer, L. et al. Assembly of high-affinity insulin receptor agonists and antagonists from peptide building blocks. *Proc Natl Acad Sci U S A* **100**, 4435-9 (2003).
54. Zhang, B. Discovery of a Small Molecule Insulin Mimetic with Antidiabetic Activity in Mice. *Science* **284**, 974-977 (1999).

55. Jensen, M., Hansen, B., De Meyts, P., Schaffer, L. & Urso, B. Activation of the insulin receptor by insulin and a synthetic peptide leads to divergent metabolic and mitogenic signaling and responses. *J Biol Chem* **282**, 35179-86 (2007).
56. Scott, P.H., Brunn, G.J., Kohn, A.D., Roth, R.A. & Lawrence, J.C. Evidence of insulin-stimulated phosphorylation and activation of the mammalian target of rapamycin mediated by a protein kinase B signaling pathway. *Proc Natl Acad Sci U S A* **95**, 7772-7777 (1998).
57. Ueki, K., Algenstaedt, P., Mauvais-Jarvis, F. & Kahn, C.R. Positive and negative regulation of phosphoinositide 3-kinase-dependent signaling pathways by three different gene products of the p85alpha regulatory subunit. *Mol Cell Biol* **20**, 8035-46 (2000).
58. Tsuruzoe, K., Emkey, R., Kriauciunas, K.M., Ueki, K. & Kahn, C.R. Insulin receptor substrate 3 (IRS-3) and IRS-4 impair IRS-1- and IRS-2-mediated signaling. *Mol Cell Biol* **21**, 26-38 (2001).
59. Ueki, K. et al. Increased insulin sensitivity in mice lacking p85beta subunit of phosphoinositide 3-kinase. *Proc Natl Acad Sci U S A* **99**, 419-24 (2002).
60. Chang, L., Chiang, S.H. & Saltiel, A.R. Insulin signaling and the regulation of glucose transport. *Molecular Medicine* **10**, 65-71 (2004).
61. Cox, M.E. et al. Insulin receptor expression by human prostate cancers. *Prostate* **69**, 33-40 (2009).
62. Duronio, V. Insulin receptor is phosphorylated in response to treatment of HepG2 cells with insulin-like growth factor I. *Biochem J* **270**, 27-32 (1990).
63. Gupta, D., Varma, S. & Khandelwal, R.L. Long-term effects of tumor necrosis factor-alpha treatment on insulin signaling pathway in HepG2 cells and HepG2 cells overexpressing constitutively active Akt/PKB. *J Cell Biochem* **100**, 593-607 (2007).
64. Varma, S. & Khandelwal, R.L. Effects of rapamycin on cell proliferation and phosphorylation of mTOR and p70(S6K) in HepG2 and HepG2 cells overexpressing constitutively active Akt/PKB. *Biochim Biophys Acta* **1770**, 71-8 (2007).
65. Bertrand, L. et al. AMPK activation restores the stimulation of glucose uptake in an in vitro model of insulin-resistant cardiomyocytes via the activation of protein kinase B. *American Journal of Physiology-Heart and Circulatory Physiology* **291**, H239-H250 (2006).
66. Hii, C.S.T., Moghadammi, N., Dunbar, A. & Ferrante, A. Activation of the phosphatidylinositol 3-kinase-Akt/protein kinase B signaling pathway in arachidonic acid-stimulated human myeloid and endothelial cells - Involvement of the ErbB receptor family. *Journal of Biological Chemistry* **276**, 27246-27255 (2001).

Chapter IV – Evaluation of Active SIRB Peptides: Binding Site, Specificity, and Potency Optimization

Substantiating the Structural Model with Indirect Evidence

At the end of the previous chapter, we have arrived at the hypothesis that SIRB peptides, much like S371 from which they are derived, not only mimic αCT of IR in sequence composition and helical structure, but also adheres to the same site –the hydrophobic patch of L1 β-sheets as the native αCT does. In order to unequivocally prove this model we would need to solve the structure of IR in complex with one of the active peptides, an effort that is currently underway. In the meantime, biochemical methods that add to the body of evidence that is internally consistent and collectively support the hypothesis are necessary efforts.

The question of how to gather such clues is not unprecedented in the field of insulin and IR research. On the contrary, this is exactly a hot area of study where the much coveted structure of ligand-bound receptor remains elusive, albeit not due to lack of attempts. Besides the fact that the insulin receptor itself is an incredibly challenging subject to express and purify[1-5], let alone crystalize (structure of the full ectodomain was not complete until 2011 at the cost of over a thousand liters of mammalian cell culture in a bio-reactor), its interaction with insulin is of such peculiar nature that further obstructs the production

of a co-crystal. In the last chapter there was a brief introduction of the two-site model speculated for insulin binding to the IR[6-8]. Each protomer of the IR homodimer purportedly contains two insulin-binding sites (hereafter referred to as sites 1 and 2 on one protomer and sites 1' and 2' on the other) that differ in affinity (figure 3.1). Binding of ligand to the low-affinity site (site 2) of one protomer is speculated to be followed by crosslinking of the half-bound ligand to the alternate site of the other protomer (site 1'). These additional contacts result in the experimentally observed high-affinity binding component, while simultaneous binding of another molecule of ligand to an unoccupied single site constitutes the experimentally observed low-affinity component. The first insulin molecule binds with high affinity by crosslinking sites 1 and 2'. On partial dissociation of the first bound ligand, a second molecule of insulin can crosslink the remaining the remaining sites 1' and 2, causing complete dissociation of the first bound insulin. Most significantly, negative cooperativity results from accelerated dissociation of the first bound ligand.

We look at what researchers have done in the absence of a structure to provide insight to the binding event that has built the two-site model. Physical evidence of the two sites themselves has come in three forms. First, alanine scan identified residues on the receptor that are hotspots of insulin binding[9,10]. Second, photo-activated crosslinking experiments covalently linked photoactive amino acid-substituted sites on insulin to residues on the receptor[11-14]. Importantly, residues implicated in both types of experiments overlap to a great degree, and locate primarily in the domains of IR discussed above, providing further confidence on the validity of binding sites. A third and final line of evidence is none other than the case in point – the phage display study that yielded IR-binding peptides, they are grouped into two classes that bind to distinct sites on the IR according to results from mutual competition assays. Although no physical data of where exactly on the receptor these peptide bind has been reported, it is strongly believed that they individually bind to

one of the two sites that insulin touches. And this last hypothesis is incidentally what we would like to prove. Interesting enough, both of the first two studies, alanine scan and photo-activated crosslinking done on the insulin-IR system can in theory be applied to the peptide-IR system as well.

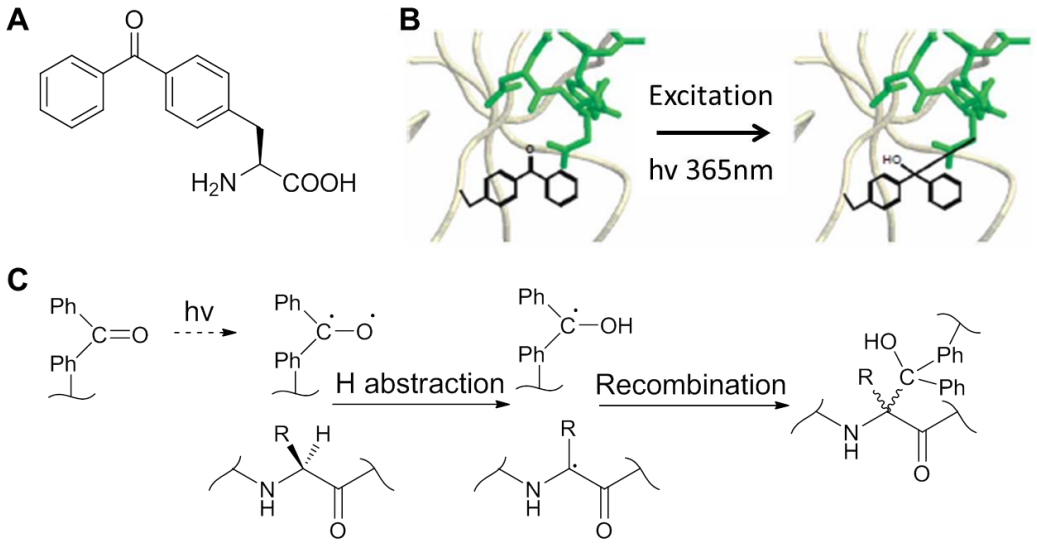


Figure 4.1. Use of photo-activated cross-linking (PCL) as a means to reveal binding site of a ligand's binding site on target protein. (A) Benzoylphenylalanine (BPA), the photo-active probe incorporated into the ligand. (B) Schematic representation of cross-linking between BPA and protein via formation of new C-C bond. (C) Reaction mechanism of PCL.

A successful study of alanine scan, as the name suggests, demands a thorough screening of all residues on the receptor suspected to be hotspots, and in this context with multiple-component binding sites means a large of number of mutants to be made. We have made the point of limited accessibility of recombinant IR (let aside production of many mutants). Photo-activated cross-linking, on the other hand, requires no perturbation on the protein but chemical modification on the peptide only. Therefore, we selected this methodology as a starting point.

Attempts at Photo-activated Cross-linking of SIRB Peptide to IR

Photo-activated cross-linking (PCL) has proved a powerful technique for probing molecular interactions[15,16]. The assay involves installment of non-natural amino acids containing various UV-responsive functional groups on a ligand that can form covalent bonds with the ligand's binding target when irradiated at the proper wavelength. The covalently linked complex is then subjected to limited proteolysis, after which the fragments can be visualized on SDS-PAGE. Oftentimes this is enough to reveal roughly where on the protein the cross-linking reaction has taken place; alternatively, the band containing the ligand is excised to be analyzed on LC-LC/MS and the exact residue on the target that is cross-linked can be identified. Although the assay has the advantage being volatile enough to adapt to almost any binding system, there are also pitfalls that could lower its success rate, such as inefficient bond formation due to low affinity, nonspecific binding and cross-linking, heterogeneity of the covalently linked complex that clouds subsequent analytical chemistry, etc.

A variety of photo-active probes, such as *para*-benzoyl-phenylalanine (BPA), *para*-azido-phenylalanine (PAP), and azetidine-based amino acids exist in the toolbox of PCL. Previous studies of the insulin: IR system have found use of both BPA[17,18] _ENREF_3 and PAP[14,19], which were substituted into a handful of residues on insulin to react with IR and yielded the important information discussed earlier. To us, these two molecules showed similar promises in working, and we chose the slightly more hydrophobic BPA to go forth for the consideration that the speculated binding site on the receptor contains a greasy patch. The reaction mechanism of BPA and a C-H bond on a protein under photo-activation and the resulting C-C bond formation is depicted by figure 4.1.

The hypothesis for our photoaffinity labeling experiment was simple: if S371 or any SIRB peptide binds to the IR consistently at a defined site (and that site is where we think it is, i.e. insulin-binding Site 1 on the receptor), then a version of the peptide derivatized with a photo-active probe should cross-link to that site upon UV irradiation on the pre-formed complex. Ensuing analytics would allow us to pinpoint the residue(s) on the IR. In execution, however, positioning of the photo-active amino acid on the peptide was not immediately obvious. On one hand, it needs to be physically close enough to the protein to react with it; on the other, it inevitably brings perturbation to the canonical binding interface that may or may not diminish binding and thus defeat the purpose. Because there was little information other than guesswork to help on decision-making, we had to empirically explore a variety of options.

As such we were again facing the need of a large synthetic project to produce the photo-active peptides, which called for narrowing down of choice of the model peptide at least. From previous screening of the SIRB library, the most potent peptide in terms of activity on the IR (antagonistic) is SIRB-B5. We reasoned that its high potency and low IC50 of IR and Akt1 phosphorylation compared to the rest of the library may arise from a higher affinity to the receptor, and therefore used its composition as a basis for making compounds for PCL.

Peptide			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
SIRB-B5-BPA0	Biotin	bAla bpa	G	S	L	D	R8	S	F	Y	D	W	F	S5	R	Q	L	G
SIRB-B5-BPA1	Biotin	bAla -	G	S	L	D	R8	S	F	Y	D	W	F	S5	bpa	Q	L	G
SIRB-B5-BPA2	Biotin	bAla -	G	S	L	D	R8	S	F	Y	D	W	bpa	S5	R	Q	L	G
SIRB-B5-BPA3	Biotin	bAla -	G	S	L	D	R8	S	F	Y	D	bpa	F	S5	R	Q	L	G
SIRB-B5-BPA4	Biotin	bAla -	G	S	L	D	R8	S	F	bpa	D	W	F	S5	R	Q	L	G
SIRB-B5-BPA5	Biotin	bAla -	G	S	L	D	R8	S	bpa	Y	D	W	F	S5	R	Q	L	G
SIRB-B5-BPA6	Biotin	bAla -	G	S	L	bpa	R8	S	F	Y	D	W	F	S5	R	Q	L	G

Figure 4.2. Incorporation of BPA into SIRB-B5. BPA was placed at either the N-terminus (SIRB-B5-BPA0) or embedded in the sequence of SIRB-B5 (SIRB-B5-BPA1-6). In the latter case, BPA replaced residues that were hypothesized to directly interact with L1 based on the structural model.

SIRB-B5 is from the *i, i*+7series, containing two non-natural amino acids used for stapling and leaving 14 other residues as well as the termini for possible sites of BPA substitution. Figure 4.2 summarizes the list of photo-active peptides designed for synthesis, termed the "SIRB-B5-BPA" series. BPA0 features the probe dangling off the N-terminus of the peptide and separated from the primary sequence by a β-Alanine linker, so as to introduce as little disruption to the original protein-protein interactions as possible. In BPA1-6, the probe is used to directly replace each of the residues at the hypothetical binding interface that actively contribute to binding via hydrophobic or electrostatic interactions (see chapter III). The hope here is that these are the "branches" that reach out farthest and closest to the protein, and should be the most promising in compensating for the apparently low binding affinity as suggested by a meek IC50.

Incorporation of BPA was quite straightforward, since Fmoc-BPA is commercially available and could basically be treated as a normal amino acid in the SPPS process (discussed in more detail in Experimental Methods). Biotin was added to the N-terminus of each peptide as a handle for pull-down or Western visualization. All seven peptides were successfully made and purified, but an alarming sign lit up as the final products were analyzed on LC/MS. Retention times of all SIRB peptides had been noted in the past to be on the long end and significantly longer than the wild-type peptide S371, plausibly a reflection of the hydrophobicity added by the staples (Series A, ones with the shortest and least greasy cross-link, fared better than the rest). Consequently, the aqueous solubility of the peptides was not impressive, but high enough to allow for experimentation being done in buffers such as PBS or cell culture media. Now with the BPA substitution the retention time climbed to a record high; some of these peptides did not elute from the C18 column until the solvent reached over 90% acetonitrile. We had but to accept the physical properties of the peptide

as they were which goes to show the intrinsic limits of the grand scheme of this investigation.

The other component of the bimolecular photo-activated reaction was the receptor. Prior to this point we had circumvented the necessity of recombinantly expressed insulin receptor, as cell-based assays served the purpose of screening the library. In order to visualize the

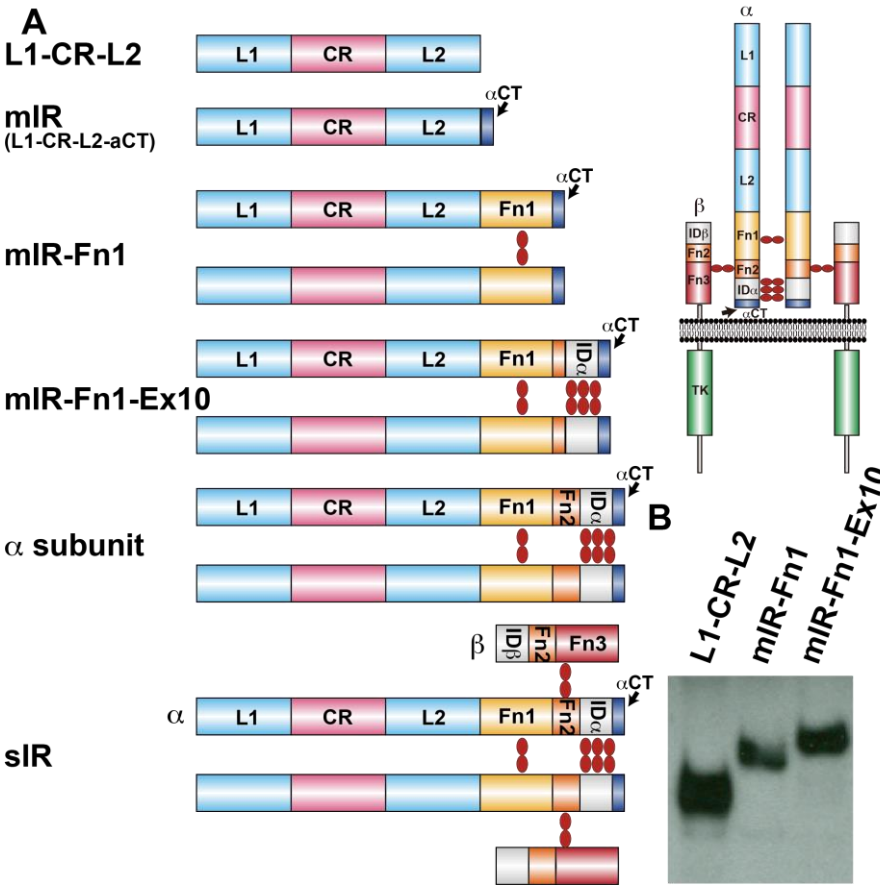


Figure 4.3. IR ectodomain constructs. (A) Truncation or deletion of the IR ectodomain that have been reported to retain high affinity for insulin will be investigated in this work. Constructs with or without α-CT will be used for crystallization and binding assays in the future with SIRB peptides in parallel to examine whether α-CT affects the interaction of SIRB peptides with the IR. (B) Expression of IR constructs. HEK293 cells were stably transfected with various constructs; protein secreted into expression medium could be visualized by Western blot of their His5 tag.

cross-linked product on SDS-PAGE or conduct downstream mass spectrometry analysis, we had to have available a reasonable amount of some form of IR ectodomain in a cell-free system. In collaboration with Dr. Minyun Zhou, a post-doctoral fellow in the lab, we designed three constructs of IR that differed in length (Figure 4.3A).

The first construct included the first three domains only and was thus termed "L1-CR-L2". This particular truncation of the α-subunit brought about the first ever crystal structure of IR[4] and, interestingly, of the type I insulin-like growth factor receptor (IGF1R)[20] before as well (no complete structure of the ectodomain of IGF1R has yet been solved to date). Referred-to as the mini-receptor or IR485 (for number of residues included), the well-studied contruct was generally thought to be easier to express, purify, and handle than any other selection of domains[2,3]. By itself L1-CR-L2 does not bind to insulin, another line of evidence the ligand-binding requires multiple components of the IR, some of which in other domains. When the α-CT segment (residues 704-719) was fused to the C-terminus of the constructs, insulin binding was reconstituted to nanomolar affinity, but no negative cooperativity characteristic of native receptor-ligand interaction was observed. Most strikingly, even the inclusion of free α-CT peptide in solution could restore the insulin-binding ability of L1-CR-L2 to a similar degree. It is remarkable how much insight on the composition of binding sites had been provided by this minimized receptor. When the long circulating belief that the β-sheets of L1 and α-CT are interaction partners *in trans* was finally proved by crystallographic data in 2010, researchers obtained another interesting result along the way[5]. Again they made use of the mini-receptor by measuring its binding affinity to the wild-type α-CT peptide and to a double mutant of α-CT on which they had substituted two residues from S371 predicted to be important for binding. They found that the K_d of the double mutant to the mini-receptor was over a thousand-fold lower than that of wild-type α-CT. Of significance was not just that a change based on phage display results

outcompeted the naturally evolved sequence (of course, one can argue that the natural α-CT might have evolved purposely NOT to be too tight a binder to L1 so as to wriggle around when insulin comes in. This was later proven to be the case; more will be discussed later), but that this was in support of the hypothesis of S371 being a structural mimic of α-CT. The corollary of the hypothesis would be the residues evolved to be important for S371's binding the receptor would exert similar increase in affinity when grafted to α-CT, the exact observation here. Combining all prior knowledge, the mini-receptor lacking native α-CT was the ideal construct for binding assays with SIRB peptides.

The other two constructs, mIR.Fn1 and mIR.Fn1-Ex10, both contain α-CT and most of the other domains of the α-subunit, and were expected to have lower affinity to SIRB peptides. mIR.Fn1 was found to bind to insulin with high affinity. mIR.Fn1-Ex10 in particular was the only incomplete combination of α-subunit domains that exhibited insulin-binding behavior identical to that of the whole ectodomain, both in terms of affinity and negative cooperativity. We envisioned that having these versions of IR which retain the natural ligand binding activity in addition to the mini-receptor would be necessary as a reality check.

All three constructs were cloned in PCDNA 3.1 expression vector with His6 and Myc tags for purification, and transfected in HEK 293 cells to generate single-colony stable cell lines that produced and secreted the protein into expression medium. Identity of each protein was confirmed by Western blot with anti-His-tag HRP-conjugated antibody (figure 4.3B). L1-CR-L2 achieved the highest expression level of all three constructs, as expected.

Purification of the protein would require a large scale of input, which took considerable time and effort to generate because of limitations of the mammalian expression system. Before the availability of pure protein, we wanted a head start on the PCL experiment with

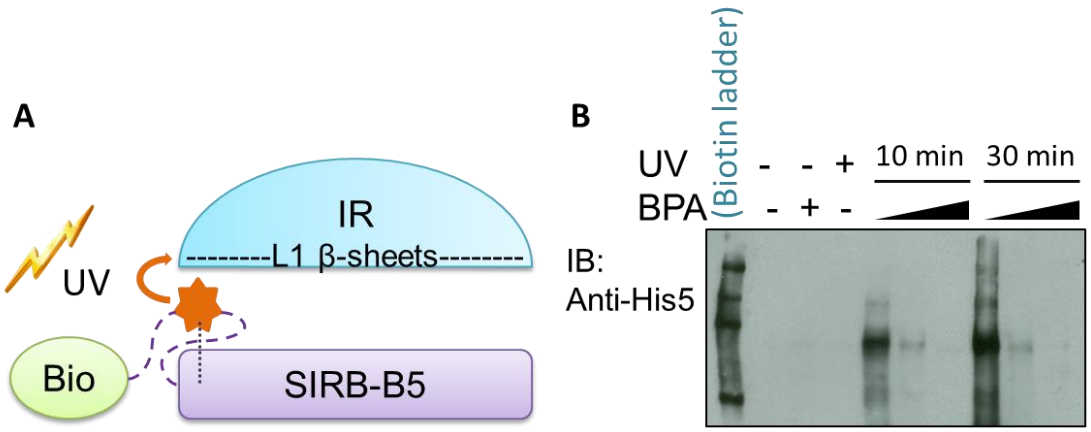


Figure 4.4. Photo-activated cross-linking of SIRB-B5-BPA and L1-CR-L2. (A) Schematic representation of PCL experiment. Biotinylated SIRB-B5-BPA allows for detection of product by Western blot. (B) Representative result (SIRB-B5-BPA0) showing lack of robustness and specificity of PCL reaction.

the expression media containing sparse but Western blot-visible amounts of protein. The assay setup is shown in figure 4.4A.

In this assay we ran into multiple obstacles. First, the aqueous solubility of BPA peptides turned out to be very low indeed – the highest concentration that they could be dissolved in the expression medium without precipitating was about 100μM. Such property was definitely against the need of having SIRB-BPA at high concentration and excess over the protein in solution to drive the complex formation and cross-linking to higher degree of completion. Second, both the efficiency and specificity of PCL were unacceptably low, manifested in smeary Western blots where a faint band of the correctly sized protein (L1-CR-L2, ~70kDa) was barely distinguishable from background (figure 4.4B). Through repeated modification of conditions and troubleshooting, the best result remained the one shown and was generated by SIRB-BPA0. With the other peptides we saw no cross-linked product whatsoever.

The PCL experiment lacked a good positive control (ideally a BPA-labeled insulin would be the gold standard, but synthesis of such a compound would be a project in its own right),

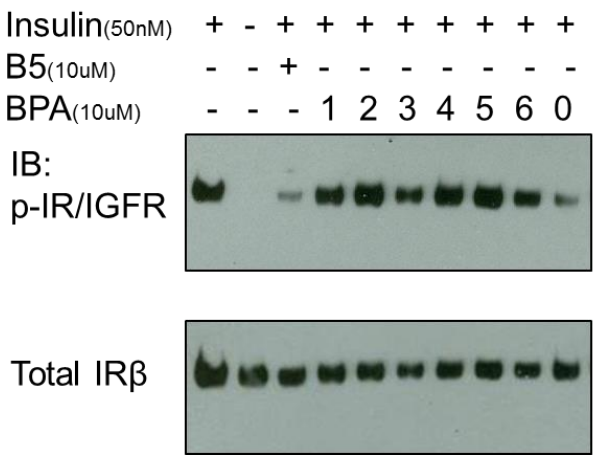


Figure 4.5. Loss of function of SIRB-B5 BPA peptides. Effects of SIRB peptides on Phospho-IR level detected by western blot. CHO-IR cells were treated with 10 μM peptides in the presence of 50 nM insulin. Antagonist effect of the original SIRB-B5 was lost in most BPA peptides.

and made it hard to diagnose where things went wrong: complex not forming? Photochemistry not happening? Light source not strong enough? To answer at least the first question, all seven SIRB-B5-BPA peptides were tested in the IR phosphorylation assay (see previous chapter). The result was as disappointing as it was illuminating: peptides BPA1-6 all significantly lost their antagonist activity to reduce insulin signaling (figure 4.5); that they would not cross-link with the receptor efficiently was hardly a surprise any longer. A likely culprit for this observed loss-of-function was, as mentioned earlier, the perturbation of the binding interface by the probe BPA. We noted that SIRB-B5-BPA0, the peptide with the probe at the N-terminus and no change in the primary sequence, did preserve the antagonist effect comparable to the original SIRB-B5, agreeing with the suspected explanation. Unfortunately as stated, even this peptide could not cross-link to any of the three constructs in a clean and satisfactory fashion, possibly because of the probe's distant location from the interface. The Western blot quality was not good enough for a concentrated sample to be collected by SDS-PAGE excision as planned.

At this point, it was clear we needed a different approach to probe the binding site of SIRB. We could launch another synthetic expedition to make more variations of the BPA peptide, or adapt other types of photo-active probe. However, the return on investment of such

endeavors was questionable since they suffered from the same trial-and-error base and lack of predictability as the previous attempt.

Despite the set-back of PCL, successful production of the three IR constructs lit up the hope for alternative means to the same end. The path we took was affinity pull-down and competition as described in the next section.

Pull-down of IR by SIRB-B5 and Competition with α-CT peptides

We took a step back and asked the question of how to establish the basic event of binding, which so far had not been supported by direct physical evidence. With the recombinant protein at hand, conditions were now fully permissible for an affinity pull-down. We remade active SIRB hits, including B5, to install biotin at their N-termini after a β-Alanine linker (referred-to as "Bio-SIRB"). Figure 4.6A depicts the experimental design of the pull-down. The biotinylated peptide could be immobilized on streptavidin-coated beads and pull down the protein. Here again the penta-His tag on the receptor made visualization very straightforward.

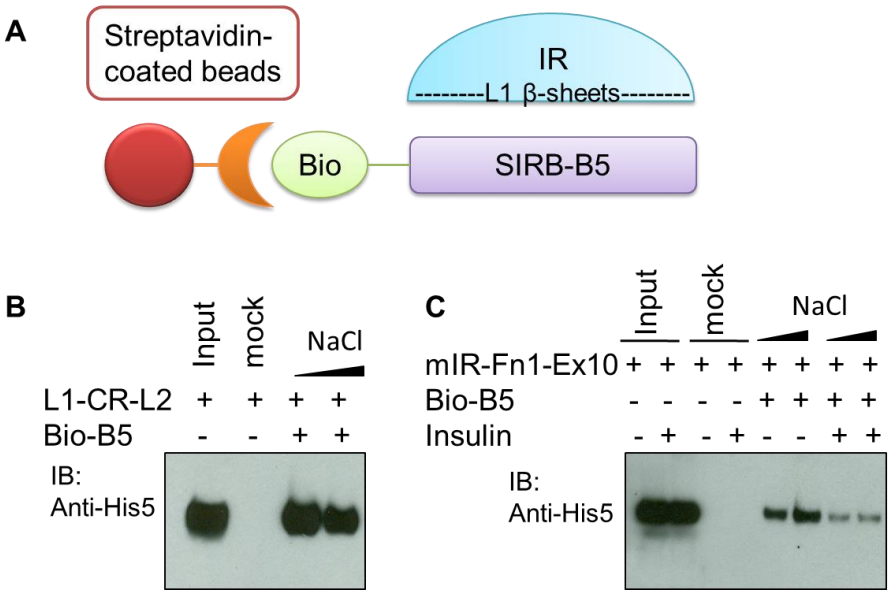


Figure 4.6. Affinity pull-down of IR by SIRB-B5. (A) Schematic representation of the pull-down assay: Biotinylated SIRB-B5 could be immobilized on streptavidin-coated beads and pull down IR from expression medium. Precipitated protein could be visualized by Western blot of His5 tag. (B) L1-CR-L2 pulled down by SIRB-B5. (C) mIR-Fn1-Ex10 pulled down by SIRB-B5 and competition by insulin.

After optimization of bead choice and binding/wash conditions, we could observe reproducible pull-down of two IR constructs, L1-CR-L2 and mIR.Fn0-Ex10 by Bio-SIRB-B5. We were confident of the specificity of binding because of 1% BSA in the binding buffer and use of a similarly-sized negative control protein that was not pulled down. Results are shown in figure 4.6B and 4.6C. Although both constructs of IR bind to Bio-SIRB-B5, L1-CR-L2 definitely does so with higher affinity, evident in higher intensity of the bands. This difference in binding affinity was as we had expected because mIR.Fn0-Ex10, unlike L1-CR-L2 , contained its native α-CT segment that could compete with SIRB-B5.

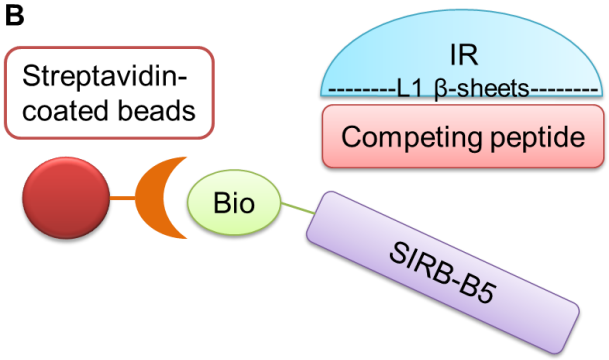
In addition to the basic pull-down of IR by SIRB, we would like to further corroborate the assay by competition with insulin. As mentioned, L1-CR-L2 loses ligand binding but mIR.Fn0-Ex10 should retain full affinity to insulin as the wild-type receptor. Therefore, titration of insulin into the complex mixture should abolish SIRB-B5 pull-down of mIR.Fn0-Ex10 at some point of concentration and have no effect on L1-CR-L2. Such indeed was the

observation (figure 4.6C). We have now gained additional confidence on the validity of the pull-down assay and established the basis of the key experiment for probing the binding site of SIRB: α-CT competition. The rationale of the experiment was that since α-CT binds to the L1 domain *in trans* is a known fact backed by both structural data biochemical studies, we could provide evidence for SIRB peptides' binding being at the same site by competing the peptide's pull-down of L1-CR-L2 with addition of free α-CT. The attractiveness of this experiment did not stop at its simplicity, but also the existence of perfect controls. We have described the double mutant (R701Y/T703W) and its increase in binding affinity to the mini-receptor earlier; this peptide could be used as a gain-of-function positive control in the experiment, i.e. a more potent competitor than the wild-type peptide. To the other end, we would like to include a loss-of-function negative mutant as well. Revisiting the structure, we

A

Peptide			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
IRalphaCT-wt	Ac	bAla	E	L	E	E	S	S	F	R	K	T	F	E	D	Y	L	H
IRalphaCT-dm	Ac	bAla	E	L	E	E	S	S	F	Y	K	W	F	E	D	Y	L	H
IRalphaCT-negmut1	Ac	bAla	E	L	E	E	S	S	A	R	K	T	A	E	D	Y	L	H
IRalphaCT-negmut2	Ac	bAla	E	L	E	E	S	S	W	R	K	T	W	E	D	Y	L	H
IRalphaCT-negmut3	Ac	bAla	E	L	E	E	S	S	D	R	K	T	D	E	D	Y	L	H
IRalphaCT-negmut4	Ac	bAla	E	L	E	E	S	S	A	R	K	T	A	E	D	A	L	H

B



C

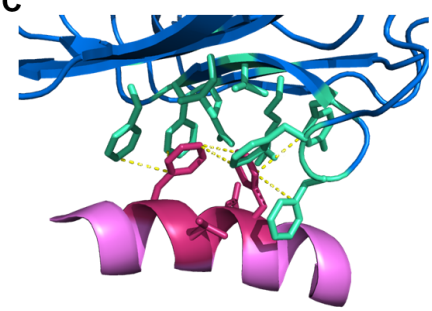


Figure 4.7. Design of competition pull-down to probe the binding site of SIRB-B5. (A) Sequences of competitive peptides: wild-type αCT, gain-of-function double mutant, and loss-of-function mutants as negative controls. (B) Schematic representation of competitive pull-down. (C) Close-up view of interaction between αCT (purple) and L1 β-sheets (blue). The two phenylalanine residues on αCT were chosen to mutate in negative control peptides.

selected the F711 and F715, two residues appear to be the most important in α-CT's binding to L1 and conserved in both IGF1R α-CT and the S371 sequence to mutate to either Ala or other disruptive residues. The peptides are summarized in figure 4.7.

As conditions of the pull-down had been optimized, we arrived at desired results at ease in the competition assay (figure 4.8.). First, negative mutants were screened, and all three mutants expectedly did not affect pull-down of L1-CR-L2 by Bio-SIRB-B5. Then α-CT_wt, α-CT_dm, and α-CT_nm1 were all titrated to the binding mixture over a range of concentrations and the results compared side-by-side. While the α-CT_nm1 did not decrease the amount of protein pulled down even at the highest concentration of 100μM, both α-CT_wt and α-CT_dm showed competition with the binding event in a dose-dependent manner. Remarkably, α-CT_dm outdid its wild-type counterpart in the competition assay by almost three orders of magnitude of concentration, quite in agreement with the its reported K_d with the mini-receptor. Barring potential allosteric changes in the receptor dimer *in situ* that were not accounted for by the truncated construct, these results would suggest that SIRB-B5 compete with α-CT for binding to the receptor,

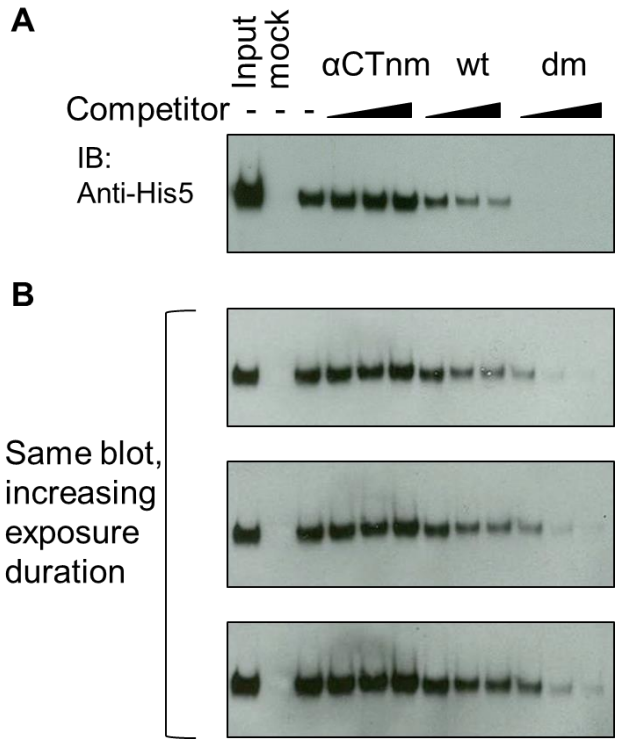


Figure 4.8. Results of competitive pull-down assay. L1-CR-L2 was pulled down by SIRB-B5 alone or in the presence of competitive peptides. (A) Concentration range of αCTnm, αCTwt, and αCTdm was 10, 50, and 100 μM. (B) Concentration range of αCTnm and αCTwt was the same as in (A), while concentration range of αCTdm was decreased to 0.1, 1, and 10 μM.

and therefore share the same binding site, i.e. the β-sheets of L1 domain.

In summary, the pull-down assay and competition with α-CT derived peptides provided indirect but concrete evidence for localization of SIRB peptides on the insulin receptor. Efforts are currently underway for obtaining structural data of the bound complex that would put the nail in the coffin.

SIRB-B5's Effect on IGF-1R

In Chapter III we investigated the effects of SIRB peptides on IR signaling, and were able to establish the on-target activity by virtue of a CHO cell line that expressed IR in large quantities but no related receptor tyrosine kinases (RTKs). However, such experiments have not ruled out possible off-target effects. Particularly of relevance is the type-1 insulin-like growth factor receptor (IGF-1R) of the IR subfamily.

IR and IGF-1R share identical gene structure and domain organizations with the exception of exon 11, which is present in the IR gene but not IGF1R. Exon 11, a 12-amino acid segment close to the C-terminus of the α-subunit, is included in one isoform of IR, IR-B, and absent in the other, IR-A. Therefore, IR-A is structurally more similar to IGF-1R and has a higher affinity to IGF-1 and IGF-2 than IR-B. Like IGF1R, IR-A is also implicated in many types of cancer. Excluding Exon 11, both isoforms of IR have high sequence similarity with IGF-1R, varying from 41% to 84% depending on the domain. Moreover, monomers of IR and IGF1R have been observed to form hybrid receptors both naturally in tissues and in tumors that express more than one member of the subfamily. Other functionally related RTKs, such as the epidermal growth factor receptor (EGFR) subfamily, have certain domains in common with IR and IGF-1R, but the overall structural homology does not come close[1,4,8,20,21].

Given such evolutionary relationships and close ties, we would expect ligand-binding and signaling pathway of the IR and IGF1R to be just as promiscuous and interweaving, which is indeed the case. Insulin and IGF-1 can both bind to IR and IGF-1R in a competitive manner. The binding affinity of ligand to non-canonical receptor is lower, but not by much: insulin and IGF-1 bind to their canonical receptor with K_d in the range of 1-2x10^{-10} M, and to the other receptor with K_d in the range of 1-2x10^{-8} M[22].

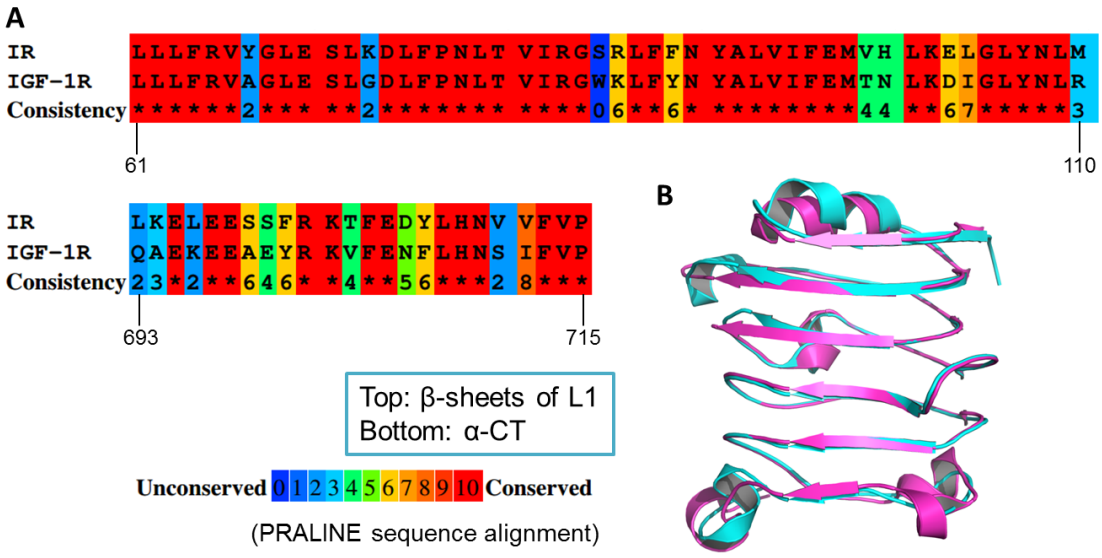


Figure 4.9. Comparison of Site 1 between IR and IGF-1R. (A) Sequence similarity of the β-sheets of L1 and αCT between IR and IGF-1R. (B) Overlay of X-ray structure coordinates of IR (cyan) and IGF-1R (purple).

All evidence suggest that S371 and SIRB peptides should bind to IGF-1R and exert similar effects on signaling, if they are indeed a partial mimic of insulin and occupy one of the multicomponent sites insulin does on both IR and IGF-1R. In fact, if we looked a bit closer at the sequence and structure of IGF-1R in both its L1 and α-CT domains, we found that the above statement is almost a corollary of the purported binding model of these peptides and has to be true. First, crystal structure encompassing the L1 domain exists for both IR and

IGF-1R, an overlay of which showed the almost identical conformations of the β-sheets (figure 4.9). Sequence homology of this particular segment in L1 reaches as high as 85% between the two receptors. Furthermore, if we focused on only the residues that interact with α-CT (based on the IR crystal structure for IGF-1R's was cut off after the first three domains), we found all but one out of 11 amino acids to be identical. The one non-homologous site is Phe vs. Tyr, still quite conserved from a structural point-of-view. Second, we examined the α-CT of the two receptors. Although the crystal structure of IGF-1R does not include its α-CT, we could take a cue from the sequence alignment with IR α-CT that shows a high degree of similarity too. All the L1-interacting residues are the same except for IR F701/Y708 and IGF-1R Y701/F708, similarly "non-different" as in the case of L1.

Besides the analysis above, we would to confirm by experiment that SIRB peptides interact with IGF-1R in a way similar to IR. The concept is quite simple – we needed to get the same kind of measurement of signaling effects on IGF-1R, i.e. Western blot or ELISA of its phosphorylation levels, but in practice the high similarity between the two receptors complicates matters. The antibody that recognizes the phosphorylated form of the receptor cannot distinguish between IR and IGF-1R. Earlier, we were able to study phospho-IR specifically thanks to the CHO-IR cell line that expressed no IGF-1R. We could try to produce a similar stable cell line expressing IGF-1R in high quantities exclusively. Alternatively, a less time-consuming route was to immuno-precipitate IGF-1R with the total IGF-1R antibody, which does discriminate against IR, before blotting for phosphorylation, as has been successfully used in a number of studies.

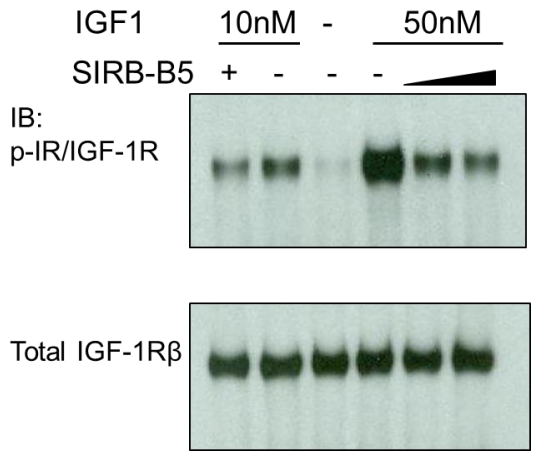


Figure 4.10. SIRB-B5 antagonism on IGF-1R. HepG2 cells were treated with SIRB-B5 in the presence of IGF1. IGF-1R was immuno-precipitated to be distinguished from IR on Phospho-IR/IGF-1R .Western blot.

HepG2 cells which have endogenous expression of both IR and IGF-1R at relatively high levels were chosen to do the experiment. We tested the antagonist effect of SIRB-B5, which had previously been thoroughly characterized on IR, on IGF-1R, with results shown in figure 4.10. Not surprisingly, SIRB-B5 demonstrates similar antagonist activity on IGF-1R, reducing its phosphorylation by IGF-1 in a dose-dependent fashion.

What are the implications of SIRB's lack of selectivity for IR vs. IGF-1R? From a therapeutic standpoint, one can argue for both the pros and cons: it renders the compounds less attractive as IR agonists used for diabetes (a lot of work needs to be done on improving the potency of agonists anyway), but raises the luring prospect of the antagonists' potential in cancer treatment as they can inhibit IGF-1R. From this dissertation's point of view, the result agreed nicely with the earlier section on the structural model of the peptides' binding to both receptors, and hence holds much appreciated value.

Covalent Dimerization of SIRB to Improve Potency

Our next attempt was to optimize SIRB hits to reach higher potency. The primary means to achieving this goal is dimerization, and here our reasoning was three-fold. First, Schaffer, L. et al. have reported use of S371 and S446, a site 2 binding peptide from the original phage display study as building blocks for peptide dimers[23]. Both were antagonists like the rest of the phage display hits (other than a few of site 2 peptides that were inactive), though S371 demonstrates weak agonistic effects at high concentrations in agreement with our preliminary screening results of series A and B (data not shown). Spacers of various lengths and different types of orientation of monomers were adapted in the synthesis to create a diverse set of homodimers and heterodimers. Of these, some were antagonists and others became agonists, including a homodimers of S371 and a heterodimer, while all had increase in binding affinity to the receptor ranging from 10- to 1000-fold compared to the monomers alone. This example clearly demonstrates the potential of producing agonistic dimers that presumably interact with both binding sites from antagonists that touch only one at a time. Second, a similar principle had been applied to other receptor tyrosine kinase systems. Phage display yielded peptides that could bind to and activate the erythropoietin (EPO) receptor but bore no sequence homology to the hormone itself[24]. When such peptides were oligomerized non-covalently, their affinity for the EPO receptor improved by 10- to 20-fold. Synthetic homodimerization of EMP1, the highest-affinity monomer and a weak agonist on its own, resulted in a 100-fold increase in affinity and an equal boost in agonistic effect of receptor activation *in vivo*. Interestingly, even an inactive peptide monomer yielded an agonistic homodimer[25]. Another noteworthy case is the discovery of a high-affinity peptide binder to the thrombopoietin (TPO) receptor and subsequent dimerization that gave rise to an agonist with a picomolar EC_{50}, which was equipotent to the natural hormone. Here the gain in activity even outweighed the gain in affinity, suggesting that improved stability due

to structural constraints could contribute directly to higher potency. Third, as we take a close look at the crystal structure of IR ectodomain, the N-termini of α-CT domains are quite close to each other (approximately 31 Å apart) across the dimer interface. Since S371 is a mimic of α-CT, it stands to reason that such an orientation further adds to the likelihood of increased affinity and potency brought by dimerization of the peptide.

In addition to the reasoning above, we noted earlier that certain SIRB peptides from the later series – D2, D5 and E2 – exhibited agonist activity at high concentrations (100μM or above). The most likely explanation was that although the peptides as monomer could not engage both Site 1 and Site 2 to induce conformational changes that activate the receptor, once their concentration is high enough for dimerization or oligomerization the non-covalent self-assembly would be able to cross-link both sites. In the earlier series, the hydrocarbon staple is short; the D and E series feature staples that are lengthier and better Van der Waals contributors, which may favor non-covalent homodimerization of the peptide. If such speculation holds, then chemically dimerizing the peptides should effectively increase the local concentration and the potency by a long way.

As for the execution of dimer synthesis, we saw three dimensions on which variations are important for arriving at an optimal compound empirically. The first dimension is the choice of peptide monomers. Guided by previous examples discussed above, we would incorporate both homodimers and heterodimers via combination of the most active SIRB peptides into this new library. Secondly, there is a variety of choice for linker (e.g. PEG or amino acid) and conjugation method (succinimide, native chemical ligation, etc.). The most important attributes of linker chemistry are arguably length and rigidity, frequently

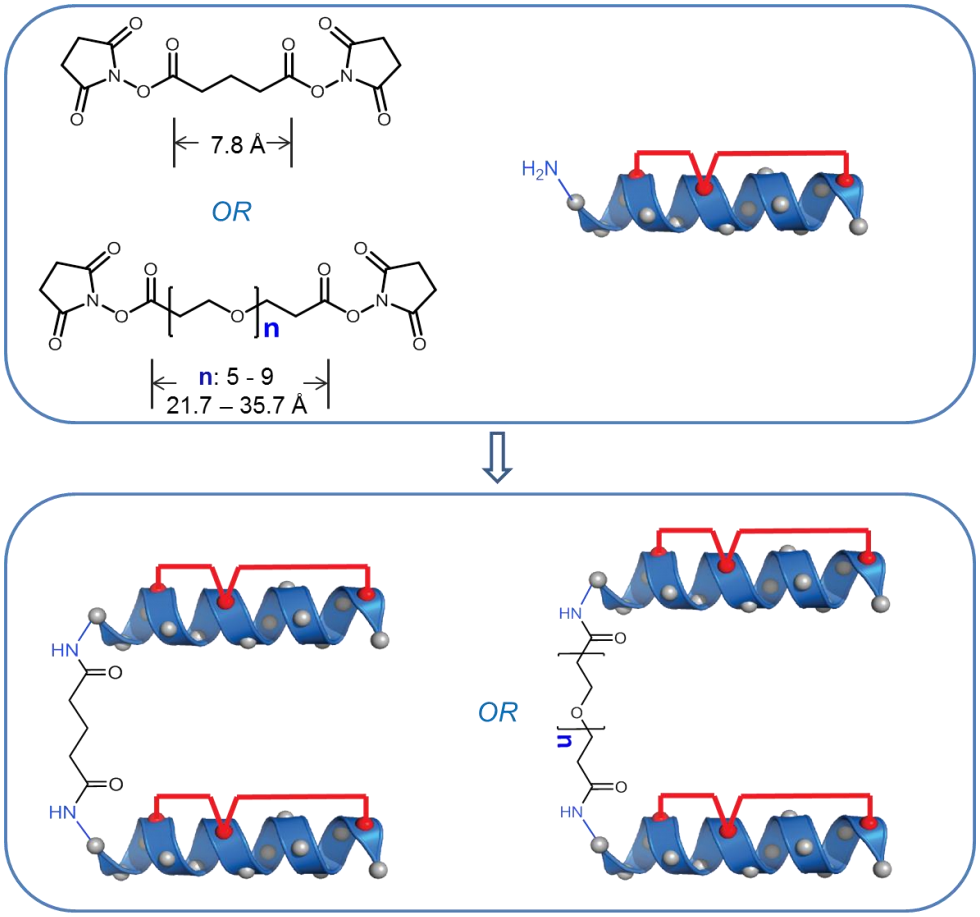


Figure 4.11. Design of SIRB homodimer synthesis. Free amine on the N-terminus of SIRB could react readily with succinimide-containing entities. Self-dimerization could be achieved in one step using a selection of bis(succinimide) linkers varying in length of spacer arm.

interdependent (shorter spacers tend to offer less conformational freedom). The last dimension of permutation is the relative orientation of the peptides – N- to N-, N- to C-, etc. Combining all three dimensions, we could see that again a diligent and collectively exhaustive synthetic project would find its worth.

While the full-blown mission has not been complete, we have made initial attempts at tackling the problem and getting preliminary data. We started with the most active antagonist, SIRB-B5, and agonist, SIRB-D2, to use as building blocks for homodimerization. As for linker chemistry, we chose a bis-succinimide system where the N-hydroxysuccinimide ester ends react readily with terminal amine of the peptide to drive homodimerization, and the spacer in between vary in length to sample a range of flexibility for the resulting dimer (figure 4.11). The three spacers were glutaric acid (Glu), PEG5, and PEG9. Such a reaction scheme constrains the orientation to be of N- to N- linkage.

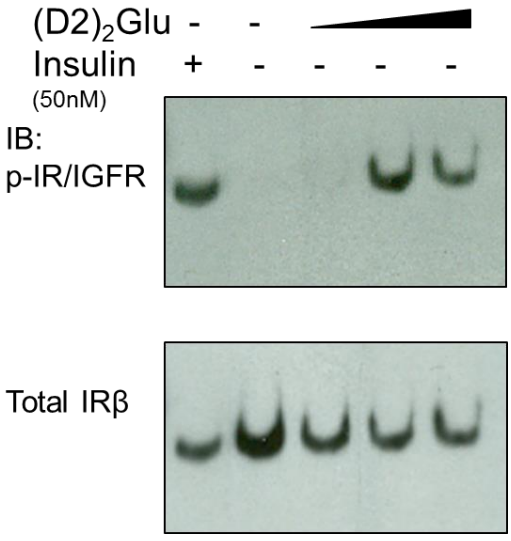


Figure 4.12. Increase of agonist potency by dimerization of SIRB-D2. CHO-IR cells were treated with increasing concentration of $(D2)_2Glu$ (1, 10, and 100 μM) in the absence of insulin. Results are shown in comparison of stimulation by 50nM insulin. Total IR levels were measured as control.

After troubleshooting the conditions for dimer synthesis, we obtained products $(B5)_2Glu$, $(B5)_2PEG5$, $(B5)_2PEG9$, and $(D2)_2Glu$, $(D2)_2PEG5$, $(D2)_2PEG9$. The dimeric peptides were subject to activity profiling in the IR and Akt1 ELISA assay as described in Chapter III. The results were very encouraging (4.12). All showed a significant shift to lower concentration in reaching maximal effects. Curiously the agonistic activity on Akt1 was much stronger than seen in IR across all compounds.

$(D2)_2Glu$ appeared to be especially active in promoting phosphorylation of both proteins. Western blot confirmed the increased potency of $(D2)_2Glu$, as at 10μM the peptide dimer could induce as much phosphorylation as 50nM insulin. This represented a 10-fold increase in potency, modest but promising for future efforts.

Topic Closing Remarks

Potency Optimization

The increase in agonist activity of SIRB-D2 homodimers compared to the monomer suggests that there is further room in affinity and potency improvement for current SIRB hits. Extensive synthesis of combinatorial homo- and heterodimers, however, must precede the hunt for the most active compounds, much like our earlier efforts on completing the first-generation SIRB library. Alternatively, obtaining structural data of current compounds bound to IR would provide instructions on rational design of better agonists and antagonists. The quest for complex crystal formation is currently underway.

In addition to dimerization of SIRB hits, there are lessons to be learned by tethering a SIRB peptide to the other class of peptides from the phage display study. For example, Novo Nordisk reported potent antagonist effects of S961, a linear combination of S371 and S661, a Site 2 peptide. In our IR phosphorylation assay, the potency of S961 was almost 1000 fold stronger than SIRB-B5. This prompts us to wonder what would happen if we extended the sequence of SIRB-B5 to include the Site 2 peptide. In other words, create a stapled version of S961. We imagine the resulting compound to retain if not exhibit even higher potency while having the advantage on pharmacokinetic properties from the staple.

Integrating New Structural Information of Partial Ligand Binding State

Recent structures of insulin in complex with IR truncations further elucidate how insulin engages its primary binding site via αCT. It has been revealed that with respect to its counterpart in apo-form IR, αCT is displaced on the L1 surface and C-terminally extended in IR-insulin complex(Figure 4.13.). Both structural and biochemical data show that the N-terminal αCT favors "antagonist conformation" without insulin binding, while C-terminal portion adopts "agonist conformation" in the presence of insulin.

As N-terminal αCT is the major determinant for L1 interactions and only C-terminal αCT is involved in insulin binding, it is reasonable to design the stapled insulin sensitizers based on C-terminus of αCT.

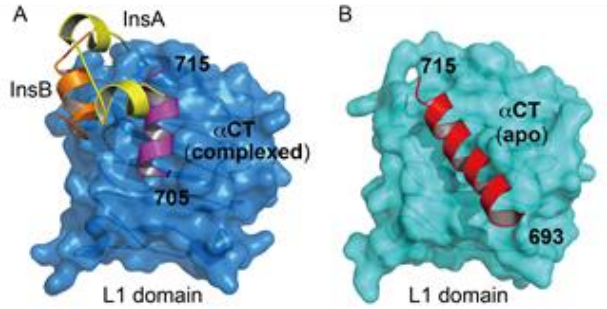


Figure 4.13. Change in position of α-CT during insulin binding. (A) α-CT (purple, 705-715) in insulin-IR complex. (B) α-CT (red, 693-715) in insulin-free IR.

Although these peptides cannot bind to IR by themselves, in the presence of insulin, they will exhibit high affinity for IR. Considering the C-terminal αCT is short (6 residues only), it allows us to introduce different elements that can either engage or block the insulin binding to the site 2 on IR and thus increase or decrease the insulin affinity for IR.

Experimental Methods

Synthesis of BPA-containing Peptides

SIRB-BPA peptides were synthesized using the standard Fmoc-SPPS method as described in Chapter I. BPA (EMD Chemicals) was treated as a "natural" amino acid and was incorporated using an amount of 3 molar equivalents to resin.

Cloning and Expression of IR Ectodomain Constructs

Full-length insulin receptor cDNA was a generous gift from Professor Joseph Bass at Northwestern University. Different truncations of the IR ectodomain, as shown in figure 4.3, were cloned into a PCDNA3.1hygro+ vector (Life Technology) using the following primers:

1) mIR-Fn1-Ex10
D2_P2_r: cagcttcagccc TGGATCCAGGGGCACAGAG
D2_P3_f: CCCCTGGATCCA gggctgaagctgccctcgag
2) mIR-Fn1
D3_P2_r: gtaatcctcaaacgt TGCCGA GGTGGCATCTGTCTGGACATAAATG
D3_P3_f: GACAGATGCCACC TCGGCA acgtttgaggattacctgcacaac
3) L1-CR-L2
D4_P2_r: gtaatcctcaaacgt CGATCC AAGTAACTCATTTTCACAGGATGCCTG
D4_P3_f: CTGTGAAAATGAGTTACTT GGATCG acgtttgaggattacctgcacaac

Resulting plasmids were transformed into DH5α competent cells and selected by ampicillin. Maxi-prepped plasmids were used to transfect HEK 293S cells using the JetPrime kit. Stably transfected cell line was generated for each protein by treatment of hygromycin for a week. Furthermore, single colony was picked from each stable cell line by dilution of original culture into to fresh expression medium and grown in 96-well plate, at a concentration such that each well would on average contain a single cell.

Expressed protein was harvested by collecting expression medium and centrifugation to remove cell debris. Expression medium was flash frozen and stored in aliquots at -20°C until use.

Pull-down Assay of IR and SIRB-B5

Preparation: Life Technology M280 magnetic streptavidin-coated beads were equilibrated in Binding Buffer and drained using the magnetic strip. Biotinylated peptides dissolved in Binding Buffer at 10μM were added to the beads, along with one sample treated with DMSO control in Binding Buffer. Immobilization of peptides was done by gentle nutating overnight in the cold room. IR construct (L1-CR-L2 or mIR_Fn1-Fn0) from expression medium was dialyzed into 1X phosphate buffered saline in the cold room overnight.

Pull-down without competition: on the next day, the beads were drained and washed thoroughly in Wash Buffer. 0.5% BSA was added to protein solution (PBS), and the solution was introduced to the peptide-bound or mock beads. Pull-down was allowed to go for up to 3h by gentle nutating in the cold room. All samples were then drained and washed extensively in Wash Buffer (either low salt or high salt). At the end of the last wash, 1X SDS Loading Buffer was added to the beads and the mixture was boiled at 95°C for at least 5 minutes to release bound protein. Samples were then run on SDS-PAGE gel to be visualized by Western blot.

Immunoblotting: At the completion of gel electrophoresis, SDS-PAGE was transferred to nitrocellulose membrane using the wet transfer method in Transfer Buffer. After transfer, the blot was blocked in Blocking Buffer (Qiagen) for up to 1h at room temperature, and incubated with anti-His5 antibody (Qiagen) at 1:1000 in Blocking Buffer overnight in the cold room. On the next day, the blot was washed extensively in tris-buffered saline, 0.1%

Tween-20 before visualization by addition of SuperSignal West Pico Chemiluminescent Substrate from Pierce.

Pull-down with Competition: the same protocol was used, with the exception that after dialysis of IR L1-CR-L2 and before addition of the protein to bead-bound peptides, the protein was incubated with indicated concentration of peptides for 1h at 4°C to allow binding.

Binding Buffer
Phosphate Buffered Saline pH 7.4 (NaCl = 147mM, low salt)
0.01%(v/v) Tween-20
0.5%(w/v) BSA

Wash Buffer
Phosphate Buffered Saline pH 7.4 (NaCl = 147mM, low salt)
0.01%(v/v) Tween-20
Phosphate Buffered Saline pH 7.4 (NaCl = 500mM, high salt)
0.01%(v/v) Tween-20

Immunoprecipitation of IGF-1R

HepG2 cells grown to 80% confluent in 10cm plates or 6-well plates were serum-starved in DMEM (penicillin/streptomycin added) overnight. On the next day, the following solutions were prepared: 50nM IGF-1 in complete growth medium (from 50μM stock, a gift from Professor Douglas Melton's group), 10μM or 100μM peptide in complete growth medium (from 10mM DMSO stock), and vehicle (the highest concentration of DMSO used in peptide samples). Old media was removed and cells were treated with appropriate compounds. The vehicle was used to create baseline, and the 50nM IGF-1 solution was used to give maximal amount of signal (100% phosphorylation). SIRB-B5 was given in combination with 50nM IGF-1. Cells were incubated at 37°C under 5% CO_2 for 15 minutes.

At the end of treatment, media was aspirated and cells were washed with ice cold PBS, and subsequently lysed using Lysis Buffer (from 10X Lysis Buffer, Cell Signaling, 1mM PMSF added prior to use). Cells were incubated in Lysis Buffer plates/wells in the cold room with gentle rocking for 10 min, then subjected to quick sonication. The cell lysate was then centrifuged at 12,000rpm for 10 min in the cold room. The clarified lysate was collected and aliquots were flash frozen in liquid nitrogen and stored at -80°C until use.

Immunoprecipitation: Clarified lysate was first quantified by the standard BCA method and all samples were adjusted to have the same total protein concentration. IGF-I Receptor β Antibody (Cell Signaling #3027) was added to the lysate at 1:100 and immunoprecipitation was allowed to proceed overnight in the cold room under gentle nutation. The next day, Protein G Magnetic Beads slurry (New England Biolabs) was added to the sample and incubated for 1h in the cold room. After extensive washing, bound protein was released from the beads and analyzed with Western blot using the same protocol as described in Chapter III.

Synthesis of SIRB homodimers

Di(N-succinimidyl) glutarate was purchased from Sigma Aldrich. BS(PEG)5 and BS(PEG)9 were purchased from Pierce. Each crosslinker was dissolved in DMSO to give a stock solution of 200mM, stored at -20°C for up to three months. SIRB peptides were synthesized as previously described, and at the end of synthesis the N-terminus was left as an unprotected primary amine.

Self-dimerization was performed by mixing 30uL of 2mM SIRB (dissolved in Reaction Buffer directly from lyophilized powder prior to reaction) and 0.25uL of 1.7mM crosslinker (from

200mM DMSO stock), and mixed gently overnight. The reaction mixture was then analyzed on LC/MS and purified on HPLC under standard procedures.

Reaction Buffer:
20% acetonitrile in PBS
pH-adjusted to pH7.5 by adding ~ 3uL 6N NaOH to a total volume of 10mL.

References

1. McKern, N.M.*, et al.* Crystallization of the first three domains of the human insulin-like growth factor-1 receptor. *Protein science : a publication of the Protein Society* **6**, 2663-2666 (1997).
2. Kristensen, C., Wiberg, F.C., Schaffer, L. & Andersen, A.S. Expression and characterization of a 70-kDa fragment of the insulin receptor that binds insulin. Minimizing ligand binding domain of the insulin receptor. *The Journal of biological chemistry* **273**, 17780-17786 (1998).
3. Kristensen, C., Andersen, A.S., Ostergaard, S., Hansen, P.H. & Brandt, J. Functional reconstitution of insulin receptor binding site from non-binding receptor fragments. *The Journal of biological chemistry* **277**, 18340-18345 (2002).
4. Lou, M.*, et al.* The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12429-12434 (2006).
5. Smith, B.J.*, et al.* Structural resolution of a tandem hormone-binding element in the insulin receptor and its implications for design of peptide agonists. *Proc Natl Acad Sci U S A* **107**, 6771-6776 (2010).
6. Ward, C.W. & Lawrence, M.C. Ligand-induced activation of the insulin receptor: a multi-step process involving structural changes in both the ligand and the receptor. *BioEssays : news and reviews in molecular, cellular and developmental biology* **31**, 422-434 (2009).
7. Ward, C.*, et al.* Structural insights into ligand-induced activation of the insulin receptor. *Acta Physiol (Oxf)* **192**, 3-9 (2008).
8. De Meyts, P. & Whittaker, J. Structural biology of insulin and IGF1 receptors: implications for drug design. *Nature reviews. Drug discovery* **1**, 769-783 (2002).
9. Mynarcik, D.C., Williams, P.F., Schaffer, L., Yu, G.Q. & Whittaker, J. Analog binding properties of insulin receptor mutants. Identification of amino acids interacting with the COOH terminus of the B-chain of the insulin molecule. *The Journal of biological chemistry* **272**, 2077-2081 (1997).
10. Mynarcik, D.C., Yu, G.Q. & Whittaker, J. Alanine-scanning mutagenesis of a C-terminal ligand binding domain of the insulin receptor alpha subunit. *The Journal of biological chemistry* **271**, 2439-2442 (1996).
11. Huang, K.*, et al.* The A-chain of insulin contacts the insert domain of the insulin receptor. Photo-cross-linking and mutagenesis of a diabetes-related crevice. *The Journal of biological chemistry* **282**, 35337-35349 (2007).
12. Kurose, T.*, et al.* Cross-linking of a B25 azidophenylalanine insulin derivative to the carboxyl-terminal region of the alpha-subunit of the insulin receptor. Identification of a new insulin-binding domain in the insulin receptor. *The Journal of biological chemistry* **269**, 29190-29197 (1994).
13. Wan, Z.*, et al.* Enhancing the activity of insulin at the receptor interface: crystal structure and photo-cross-linking of A8 analogues. *Biochemistry* **43**, 16119-16133 (2004).
14. Wan, Z.L.*, et al.* Diabetes-associated mutations in human insulin: crystal structure and photo-cross-linking studies of a-chain variant insulin Wakayama. *Biochemistry* **44**, 5000-5016 (2005).

15. Kawamura, A. & Mihai, D.M. Probing proteomes with benzophenone photoprobes. *Methods Mol Biol* **803**, 65-75 (2012).
16. Prestwich, G.D., Dorman, G., Elliott, J.T., Marecak, D.M. & Chaudhary, A. Benzophenone photoprobes for phosphoinositides, peptides and drugs. *Photochemistry and photobiology* **65**, 222-234 (1997).
17. Backer, J.M., Shoelson, S.E., Haring, E. & White, M.F. Insulin receptors internalize by a rapid, saturable pathway requiring receptor autophosphorylation and an intact juxtamembrane region. *The Journal of cell biology* **115**, 1535-1545 (1991).
18. Shoelson, S.E., Lee, J., Lynch, C.S., Backer, J.M. & Pilch, P.F. BpaB25 insulins. Photoactivatable analogues that quantitatively cross-link, radiolabel, and activate the insulin receptor. *The Journal of biological chemistry* **268**, 4085-4091 (1993).
19. Huang, K.*, et al.* How insulin binds: the B-chain alpha-helix contacts the L1 beta-helix of the insulin receptor. *Journal of molecular biology* **341**, 529-550 (2004).
20. Garrett, T.P.*, et al.* Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. *Nature* **394**, 395-399 (1998).
21. Brierley, G.V.*, et al.* Silencing of the insulin receptor isoform A favors formation of type 1 insulin-like growth factor receptor (IGF-IR) homodimers and enhances ligand-induced IGF-IR activation and viability of human colon carcinoma cells. *Endocrinology* **151**, 1418-1427 (2010).
22. Schumacher, R., Mosthaf, L., Schlessinger, J., Brandenburg, D. & Ullrich, A. Insulin and insulin-like growth factor-1 binding specificity is determined by distinct regions of their cognate receptors. *The Journal of biological chemistry* **266**, 19288-19295 (1991).
23. Schaffer, L.*, et al.* Assembly of high-affinity insulin receptor agonists and antagonists from peptide building blocks. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4435-4439 (2003).
24. Wrighton, N.C.*, et al.* Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* **273**, 458-464 (1996).
25. Wrighton, N.C.*, et al.* Increased potency of an erythropoietin peptide mimetic through covalent dimerization. *Nature biotechnology* **15**, 1261-1265 (1997).